

UNIVERSIDADE DE LISBOA  
FACULDADE DE CIÊNCIAS  
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL



## **Análise Estatística das Reclamações da EDP Distribuição: o caso particular das “Visitas Não Realizadas”**

Inês Anselmo Graça

**Mestrado em Matemática Aplicada à Economia e Gestão**

Versão Pública

Trabalho de Projeto orientado por:  
Professora Doutora Maria Isabel Fraga Alves  
Engenheiro João Filipe Neto Viegas Nunes



## Agradecimentos

Gostaria de dedicar este espaço a diversas pessoas que foram fundamentais para a realização deste trabalho.

Primeiramente, à minha família por terem acreditado sempre em mim, em especial, à minha mãe por todo o esforço e apoio durante o meu percurso académico. Por todos os fins-de-semana que foram fundamentais para restabelecer energias.

À minha orientadora, professora Maria Isabel Fraga Alves, por toda a disponibilidade, paciência, apoio e partilha de conhecimentos.

À minha equipa na EDP Distribuição por me terem acolhido tão bem e por toda a aprendizagem que me transmitiram, em particular, ao meu chefe João Nunes, pelos votos de confiança e oportunidades oferecidas. Foi um prazer ter trabalho e crescido convosco.

Às minhas companheiras de treinos por me ajudarem a começar todos os dias com um sorriso na cara e com uma outra motivação.

Ao Márcio por estar sempre lá. Por todas as palavras e apoio constantes.

Por fim, um agradecimento muito especial ao meu namorado, Miguel, por acreditar sempre em mim. Por ter lidado com as minhas inseguranças, por toda ajuda, paciência e compreensão que foram essenciais para a realização deste trabalho.

Um grande **obrigado** a todos!



*“Your most unhappy customers are your greatest source of learning”*  
– Bill Gates (1955 - ), *Microsoft*



## Resumo

Num mundo empresarial cada vez mais competitivo, onde as tecnologias têm vindo a ganhar um destaque essencial e também onde os produtos/serviços representam cada vez mais um papel banal para a angariação e/ou fidelidade dos Clientes, é necessário recorrer a diferentes técnicas que permitem às empresas diferenciarem-se da concorrência.

Uma dessas técnicas passa por tratar e analisar as reclamações das empresas e, deste modo, perceber o que o Cliente procura num produto/serviço e quais os principais pontos críticos a melhorar.

Assim, o presente estudo pretendeu fazer uma avaliação das reclamações relacionadas com os agendamentos da EDP Distribuição, mais concretamente, as “visitas não realizadas”. Para esse efeito, inicialmente foram introduzidos os conceitos de satisfação do Cliente; as diferentes formas de recolher o seu *feedback*; quais os motivos que o levam a reclamar e ainda as vantagens que uma reclamação pode trazer para as empresas. Foi também feita uma breve apresentação do setor elétrico em Portugal, assim como o papel da EDP Distribuição.

Como concretização dos conceitos ilustrados na parte teórica, recorreu-se à inferência estatística através de uma metodologia qualitativa - testes não paramétricos -, sendo a população em estudo os Clientes da EDP Distribuição no ano de 2017.

Os resultados desta análise evidenciaram a existência de variáveis que influenciam os números das reclamações, tal como sugerido na bibliografia. Neste caso, tanto a zona do país como a época do ano foram duas variáveis que apresentaram uma certa dependência em relação às reclamações de visitas não realizadas.

Para além desta análise, também se procedeu a uma comparação entre esta tipologia de reclamação e alguns dados sociodemográficos da população portuguesa, tendo por base um estudo realizado por Metehan *et al* (2011), com o intuito de averiguar a existência de alguma relação entre estas variáveis. Assim, concluiu-se que há indícios para afirmar que variáveis tais como a idade, o poder de compra e as habilitações literárias podem apresentar uma dependência com a atitude para reclamar.

Por fim, são apresentadas algumas das limitações deste trabalho, tal como algumas sugestões para trabalhos futuros dentro deste âmbito das reclamações da EDP Distribuição.

**Palavras-chave:** Experiência do Cliente, satisfação, reclamações, inferência estatística, métodos não paramétricos

## Abstract

Technologies are becoming more and more essential and goods/services have an increasingly ordinary role for Customer acquisition/fidelization in an increasingly competitive business world. Thus, using different techniques is necessary to allow companies to differentiate themselves from their competition.

One of such techniques consists of processing and analyzing data from Customer complaints gathered by the companies to understand what a Customer is looking for in a product/service and the critical flaws needed to be improved.

This study looked forward to evaluating the complaints related to EDP Distribuição's scheduling, mainly, scheduled and "unrealized visits". Initially, to complete that evaluation, the concepts of Customer satisfaction, different types of client feedback collection, reasons behind the complaints and what are the advantages that a complaint can bring to a company were introduced. It was also included a brief presentation of the energy sector along with EDP Distribuição's role.

To sustain the theory, it was used a qualitative methodology, more specifically statistic inference - nonparametric tests - considering EDP Distribuição's 2017 customers as the sample scope of analysis.

The study's results show that, as suggested in the bibliography, there are variables that influence the number of complaints, in this case, both the location and the time of the year evidenced a certain dependency related with the unrealized visits' complaints.

It was also possible to make an analysis of the relation between the complaints and certain sociodemographic data from the Portuguese population to sustain the used bibliography, mainly the study carried out by Metehan *et al* (2011). Therefore, it was concluded that there is proof that variables such as age, purchasing power and educational qualifications may show a dependency with the Customer's attitude when making a complaint.

Finally, this study's limitations are shown, together with suggestions for future papers related to the subject of EDP Distribuição's complaints.

**Keywords:** Customer experience, satisfaction, complaints, statistical inference, nonparametric methods



# Índice

|  |      |
|--|------|
| ÍNDICE DE FIGURAS .....                                    | VIII |
| ÍNDICE DE TABELAS .....                                    | IX   |
| SIGLAS E ACRÓNIMOS.....                                    | X    |
| NOTAÇÕES.....  | XI   |
| INTRODUÇÃO.....  | 1    |
| PARTE I .....  | 2    |
| 1. EXPERIÊNCIA DO CLIENTE .....                            | 3    |
| 1.1 Introdução .....                                       | 3    |
| 1.2 A Qualidade de Serviço e a Satisfação do Cliente ..... | 3    |
| 1.3 Reclamações .....                                      | 5    |
| 1.3.1 Gestão das Reclamações.....                          | 7    |
| 2. O NEGÓCIO .....   | 9    |
| 2.1 O Setor Elétrico em Portugal .....                     | 9    |
| 2.2 O Grupo EDP .....                                      | 10   |
| 2.2.1 A Visão.....   | 10   |
| 2.2.2 Evolução Histórica .....                             | 11   |
| 2.3 EDP Distribuição .....                                 | 12   |
| 3. METODOLOGIAS.....                                       | 14   |
| 3.1 Introdução .....                                       | 14   |
| 3.2 Inferência Estatística.....                            | 15   |
| 3.2.1 Teste para uma Proporção .....                       | 19   |
| 3.2.2 Teste para a Diferença de Proporções .....           | 21   |
| 3.2.3 Tabelas de Contingência.....                         | 23   |
| 3.2.3 Teste do Qui-Quadrado .....                          | 24   |
| 3.2.3.1 Pearson ( $\chi^2$ ) .....                         | 24   |
| 3.2.3.1.1 Teste Qui-Quadrado para Homogeneidade.....       | 25   |
| 3.2.3.1.2 Teste Qui-Quadrado para Independência.....       | 26   |
| 3.2.3.2 Razão de Verossimilhança ( $G^2$ ).....            | 27   |
| 3.2.4 <i>Odds Ratio</i> .....                              | 28   |
| 3.2.5 Comparações Múltiplas .....                          | 29   |
| 3.2.5.1 Procedimento de Bonferroni .....                   | 30   |
| 3.2.5.2 Procedimento de Holm.....                          | 31   |
| 3.2.5.3 Procedimento de Marascuilo .....                   | 31   |
| 3.2.5.4 Análise de Resíduos .....                          | 32   |
| BIBLIOGRAFIA .....   | 34   |

# Índice de Figuras

## Parte I

|   |    |
|---|----|
| <b>Figura 1.1</b> – Qualidade de serviço vs. expectativas .....   | 4  |
| <b>Figura 1.2</b> – Modelo do processo de decisão para reclamar.....  | 7  |
| <b>Figura 1.3</b> – Motivos das reclamações .....   | 7  |
| <b>Figura 1.4</b> – Processo de gestão da reclamação com base na justiça .....                                | 8  |
| <b>Figura 2.1</b> – Descrição da cadeia de valor do setor elétrico.....                                       | 9  |
| <b>Figura 2.2</b> – História do Grupo EDP.....  | 11 |
| <b>Figura 2.3</b> – Tipos de atividades que constituem o negócio de distribuição de energia elétrica .....    | 12 |
| <b>Figura 2.4</b> – Direções de Redes e Concessões .....  | 13 |
| <b>Figura 3.1</b> – Tipos de variáveis .....  | 15 |
| <b>Figura 3.2</b> – <i>P-value</i> para as hipóteses $H_0: \pi_1 = \pi_2$ vs. $H_1: \pi_1 \neq \pi_2$ . ..... | 22 |
| <b>Figura 3.3</b> – <i>P-value</i> para as hipóteses $H_0: \pi_1 \leq \pi_2$ vs. $H_1: \pi_1 > \pi_2$ . ..... | 22 |
| <b>Figura 3.4</b> – <i>P-value</i> para as hipóteses $H_0: \pi_1 \geq \pi_2$ vs. $H_1: \pi_1 < \pi_2$ .....   | 22 |

# Índice de Tabelas

## Parte I

|  |    |
|--|----|
| <b>Tabela 3.1</b> – Testes não paramétricos: vantagens e desvantagens.....     | 17 |
| <b>Tabela 3.2</b> – Tabela de contingência de dupla entrada $I \times J$ ..... | 23 |

## Siglas e Acrónimos

|        |  |
|--------|--|
| ARC    | Área de Redes e Clientes                     |
| B2B    | <i>Business-to-business</i>                  |
| CEE    | Comunidade Económica Europeia                |
| CPE    | Companhia Portuguesa de Eletricidade         |
| EDP    | Energias de Portugal, S.A.                   |
| ERSE   | Entidade Reguladora dos Serviços Energéticos |
| DRCN   | Direção de Redes e Concessões do Norte       |
| DRCP   | Direção de Redes e Concessões do Porto       |
| DRCM   | Direção de Redes e Concessões do Mondego     |
| DRCT   | Direção de Redes e Concessões do Tejo        |
| DRCL   | Direção de Redes e Concessões de Lisboa      |
| DRCS   | Direção de Redes e Concessões do Sul         |
| FDR    | <i>False Discovery Rate</i>                  |
| FWER   | <i>Family-Wise Error Rate</i>                |
| i.i.d. | Independentes e Identicamente distribuídas   |
| NPS    | <i>Net Promoter Score</i>                    |
| NUTS   | Nomenclaturas das Unidades Territoriais      |
| ORD    | Operador de Redes de Distribuição            |
| PSE    | Prestadores de Serviços Externos             |
| REN    | Rede Elétrica Nacional                       |
| RQS    | Regulamento de Qualidade de Serviço          |
| S.E.   | Erro padrão ( <i>Standard Error</i> )        |
| SIEG   | Serviço de Interesse Económico Geral         |
| VoC    | <i>Voice of Customer</i>                     |
| VNR    | Visitas Não Realizadas                       |
| vs.    | <i>versus</i>                                |

## Notações

|   |  |
|---|--|
| $H_0$   | Hipótese nula  |
| $H_1$   | Hipótese alternativa   |
| $\alpha$                                      | Nível de significância (erro tipo I)   |
| $\beta$                                       | Erro tipo II   |
| $\pi$   | Probabilidade de “sucesso” / Proporção populacional                              |
| $\hat{\pi}$                                   | Estimador de $\pi$ / Proporção amostral  |
| $\tilde{y}$                                   | Amostra dos valores observados de $Y$  |
| $L(\pi; \tilde{y})$                           | Verosimilhança de $\tilde{y}$ e do parâmetro desconhecido $\pi$                  |
| $\tilde{X}$                                   | Amostra recolhida de uma população $X$   |
| $S(\tilde{X}; \pi)$                           | Variável aleatória que representa uma função de $\tilde{X}$ e do parâmetro $\pi$ |
| $\bar{X}$                                     | Média de $X$   |
| $q_{1-\alpha}$                                | Quantil de probabilidade $1-\alpha$ da distribuição normal (0,1)                 |
| $\Phi(x)$                                     | Função da distribuição normal-padrão no ponto $x$                                |
| $X_n \xrightarrow[n \rightarrow \infty]{d} X$ | Convergência em distribuição   |
| $N(\mu, \sigma^2)$                            | Distribuição normal de valor médio $\mu$ e variância $\sigma^2$                  |
| $X_n \xrightarrow[n \rightarrow \infty]{p} X$ | Convergência em probabilidade  |
| $\pi_{ij}$                                    | Probabilidade conjunta   |
| $\pi_{j i}$                                   | Probabilidade condicional  |
| $\chi_n^2$                                    | Distribuição qui-quadrado com $n$ graus de liberdade                             |
| $e_{ij}$                                      | Frequências esperadas  |
| $\hat{e}_{ij}$                                | Estimadores das frequências esperadas  |
| $\Lambda$                                     | Razão de verosimilhança  |
| $\Omega$                                      | <i>Odd</i>   |
| $\theta$                                      | <i>Odds Ratio</i>  |
| $\hat{\theta}$                                | <i>Odds Ratio</i> amostral   |
| $s_{ij}$                                      | Valores críticos (procedimento de Marascuilo)                                    |
| $r_{ij}$                                      | Resíduos padronizados  |

# Introdução

O crescente enfoque na qualidade dos serviços e na satisfação do Cliente é um fenómeno relativamente recente nas organizações que tem servido como vantagem competitiva face à concorrência.

Deste modo, as organizações devem dispor de diversos sistemas de forma a recolher o *feedback* dos seus Clientes e assim ter uma avaliação da qualidade dos serviços, obtendo também informação útil para a implementação de melhorias nos serviços prestados.

Uma das melhores formas de recolher esta opinião dos Clientes é através da análise das reclamações feitas à empresa. Estas têm vindo a ganhar uma maior importância com o decorrer do tempo, conforme verificado na bibliografia, sendo fundamental para as empresas uma boa gestão das reclamações, garantido não só que tomam conhecimento dos problemas, como também adquirindo uma nova oportunidade de reter os Clientes (se aplicável).

Tendo em consideração o descrito anteriormente, para este trabalho optou-se por fazer uma análise às reclamações da EDP Distribuição (mediante autorização da mesma), com foco nas reclamações relacionadas com visitas não realizadas (VNR), referentes ao ano 2017, recorrendo à estatística não paramétrica.

De modo a concretizar o objetivo deste estudo foram distinguidas diferentes etapas:

- 1) Verificar se existiu uma evolução das reclamações de VNR em 2017 face ao ano anterior através de um teste para as proporções;
- 2) Determinar as variáveis que têm impacto nas reclamações de VNR, recorrendo aos diferentes testes do qui-quadrado;
- 3) Avaliar quais as categorias, dentro de cada variável, que podem provocar um maior impacto nas reclamações, a partir dos resíduos padronizados e também dos testes de comparações múltiplas com os respetivos ajustamentos para os *p-values*;
- 4) Compreender se existe uma relação entre os diversos fatores sociodemográficos apresentados na bibliografia e o número das reclamações de VNR.

Para efetuar o tratamento estatístico dos dados recorreu-se ao *software* R (versão 3.4.3).

Assim, de forma a cumprir as diferentes etapas propostas, na primeira parte deste trabalho é apresentada uma revisão da literatura separada em três capítulos:

No capítulo 1 são introduzidos os conceitos relacionados com a experiência e satisfação do Cliente e a importância da gestão de reclamações para qualquer negócio, como também estão mencionados alguns dos trabalhos que serviram de impulso para o estudo da relação entre as reclamações e os dados sociodemográficos; No capítulo 2 é introduzida a área de negócio em que este estudo se foca, explicando como funciona o setor elétrico em Portugal e o papel da EDP Distribuição na cadeia de valor deste setor; Por fim, no capítulo 3 é descrita a metodologia a utilizar, que contempla os diversos testes que servem de base a este trabalho.

Na segunda parte são descritos e discutidos os resultados, onde se apresentam as principais variáveis que (e como) influenciam este tipo de reclamações. Assim como é também elaborada uma breve comparação entre as reclamações e os diferentes dados sociodemográficos de Portugal continental, sustentada pelo estudo feito por Metehan *et al* (2011), dando a sugestão de quais os fatores sociodemográficos que podem ter um impacto nas reclamações.

## Parte I

---

# 1. Experiência do Cliente

## 1.1 Introdução

Atualmente, as empresas vivem num ambiente altamente competitivo, onde cada Cliente além de mais ativo, apresenta novas exigências, pelo que um ponto menos positivo em qualquer contacto pode resultar na sua insatisfação e, na pior das situações, o abandono do Cliente para uma empresa concorrente.

Para colmatar esta nova posição dos Clientes face às empresas, não só é necessário garantir a qualidade dos produtos (e serviços), como também conhecer os seus consumidores, percebendo as suas necessidades e o que procuram, melhorando assim a experiência com a empresa e, consequentemente, trazendo uma vantagem competitiva.

Como tal, a experiência do Cliente ao longo dos anos tem vindo a ganhar uma grande importância para as empresas. Um estudo realizado pela *Accenture Interactive* (2015) aferiu que “melhorar a experiência do Cliente” está no topo da lista de prioridades das empresas para os 12 meses seguintes e, seguidamente, encontram-se o “crescimento das receitas” e a “diferenciação”. Também Drucker (1993), citado em Battaglia *et al* (2010), afirma que o objetivo de uma empresa comercial não é “fazer dinheiro”, mas sim criar o Cliente e satisfazê-lo, sendo a sua satisfação crucial para a sobrevivência de qualquer organização.

## 1.2 A Qualidade de Serviço e a Satisfação do Cliente

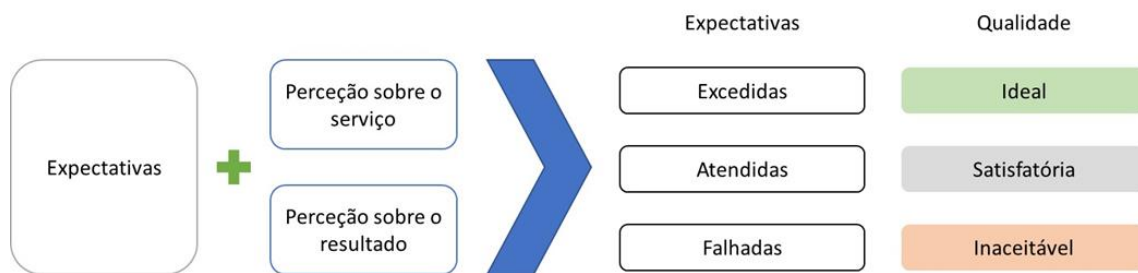
Desde os anos 80 que a qualidade dos bens e serviços se tem tornado uma preocupação para as organizações, sendo que a qualidade dos produtos pode ser facilmente avaliada, mas o mesmo não pode ser assegurado em relação à qualidade dos serviços, uma vez que corresponde ao tratamento de um Cliente ou a algo que lhe pertença (Parasuraman, 1985).

Mas, a qualidade nem sempre é o suficiente para garantir a satisfação dos Clientes. Fornell *et al* (1987) afirmam que não é possível por parte de uma empresa garantir a todo o momento 100% de satisfação de todos os seus Clientes devido a diversas causas, como por exemplo, o desejo pela variedade. Porém, o custo de novas angariações é superior em cerca de cinco vezes em relação ao custo de reter os atuais Clientes (Invesp, 2017). Assim, é importante para as empresas compreenderem os seus erros e como podem melhorar a experiência dos Clientes. A título de exemplo, em “*The CEO Guide to Customer Experience*” (McKinsey & Company, 2016) foi criado um modelo com o objetivo de melhorar a experiência do Cliente, onde um dos passos mencionados corresponde à observação “*pelo olhar do Cliente*”, de modo a compreender quais as suas necessidades e as suas expectativas.

A relação “empresa-Cliente” é estabelecida através do contacto entre o consumidor e o prestador de serviços. Cada contacto é denominado por “momento da verdade” e é neste momento que o Cliente forma a imagem da empresa, sendo um elemento crucial para a satisfação do mesmo com a empresa.

Segundo Berry *et al* (1985) as expectativas criadas pelo Cliente em relação ao serviço, representam um ponto-chave na avaliação do mesmo, uma vez que o Cliente compara as suas expectativas com a experiência obtida, retirando assim conclusões em relação à sua qualidade (ver figura 1.1).





**Figura 1.1** - Qualidade de serviço vs. expectativas  
Fonte: Adaptado de Berry *et al* (1985)

De acordo com Parasuraman *et al* (1985), a falta de correspondência entre as expectativas criadas e o serviço prestado pode ser consequência de pelo menos um dos cinco *gaps*:

- i) Discrepância entre as expectativas do Cliente e a percepção das mesmas pela empresa;
- ii) Discrepância entre o entendimento da empresa e as especificações que elabora para atender o Cliente. (Por exemplo, um serviço mal interpretado);
- iii) Discrepância entre o serviço que se pretende produzir e o serviço que é efetivamente realizado. (Por exemplo, falta de conhecimento da parte dos prestadores de serviço);
- iv) Discrepância entre a qualidade especificada do serviço e a comunicação externa da empresa. (Por exemplo, divergência entre o que é prometido em campanhas publicitárias e o que é realmente vendido);
- v) Discrepância entre as expectativas do Cliente sobre o serviço e o serviço efetivamente prestado. Resultado da junção dos 4 *gaps* anteriores.

Porém, existem diversos fatores que podem influenciar estas expectativas, tais como: o *Word-of-Mouth* (ou o “passa-a-palavra”), onde o potencial Cliente recebe opiniões de outros Clientes; as necessidades pessoais, que correspondem à natureza do serviço e à sua urgência; e por fim, as experiências anteriores do Cliente com a empresa (Parasuraman *et al*, 1985).

Uma forma das empresas perceberem como podem melhorar a sua relação com os Clientes é através dos seus *feedbacks*, utilizando essa informação para melhorar e implementar novas medidas, se necessário. Para registrar o *feedback*, as empresas podem recorrer a abordagens tais como a escala Servqual, o *Net Promoter Score (NPS)*, entre outras.

A escala Servqual foi criada por Parasuraman, Zeithaml e Berry, em 1988, com o intuito de medir a qualidade do serviço sob a ótica dos Clientes, com base nas suas percepções, após a realização do mesmo. Esta escala foi construída com base nos cinco *gaps* apresentados anteriormente e pretende que o Cliente atribua uma pontuação relativa à qualidade dos serviços e consoante aquilo que considera importante num serviço (Parasuraman *et al*, 1988).

Por outro lado, o *NPS* é uma metodologia criada por Fred Reichheld que consiste numa mensuração do tratamento dos Clientes por parte da empresa, avaliando a sua lealdade para com a mesma. Sendo esta bastante frequente em grandes nomes no ramo empresarial, tais como *Apple*, *Allianz* ou *EDP*, devido à sua simplicidade, flexibilidade e também confiabilidade (Reichheld *et al*, 2011).

Porém, existe uma alternativa a este tipo de metodologias que também permite uma recolha das opiniões dos Clientes, conhecendo as suas principais insatisfações e os pontos que precisam de ser aprimorados – as reclamações.

### 1.3 Reclamações

Em Portugal, todos os fornecedores de bens e prestadores de serviços que exercem atividade em “estabelecimentos de venda ao público e de prestação de serviços de comércio a retalho e conjuntos comerciais, bem como estabelecimentos de comércio por grosso com revenda ao consumidor final” ou em lojas online são obrigados<sup>1</sup> a ter e disponibilizar o livro de reclamações, fazendo com que todas as empresas destas atividades tenham em sua posse uma forma de recolher as principais insatisfações dos seus Clientes.

Geralmente, as reclamações são vistas como algo negativo para as empresas, uma vez que refletem o descontentamento dos seus consumidores. Contudo, estas devem ser interpretadas como uma oportunidade de recolher os seus *feedbacks*, ajudando a perceber a origem das suas insatisfações e como resolver os problemas (Tax *et al*, 1998).

Diversos autores (Spreng, 1995; Battaglia *et al*, 2010) afirmam que muitas vezes os Clientes que não reclamam, optam por abandonar a empresa, pelo que se deve encorajar os mesmos a reclamar à organização que existem deficiências que devem ser corrigidas, oferecendo assim uma hipótese de se restabelecer a satisfação.

Por outro lado, alguns estudos, citados em Boote (1998), mostram que apenas um em cada cinco Clientes insatisfeitos reclamam à organização, levando assim a concluir que existem outras motivações externas que podem influenciar um consumidor a reclamar, tais como dados demográficos, culturais ou fatores sociais, por exemplo.

Em Metehan *et al* (2011) são mencionados vários autores que se dedicaram à análise desta relação entre dados sociodemográficos e a atitude para com a reclamação. Este estudo serviu também de base para uma das análises apresentada nesta dissertação, onde se pretende comparar a relação das reclamações com algumas das variáveis sociodemográficas mencionadas, tais como:

- Género

Esta é uma das variáveis que tem gerado uma maior controvérsia relativamente à existência de relação com as reclamações. De acordo com alguns autores (Liefeld *et al*, 1975; Phau and Baird, 2008) não existe qualquer diferença entre os géneros, porém autores como Heung and Lam (2003) e Kau *et al* (1995) sugerem que os homens reclamam mais do que as mulheres.

- Idade

Metehan *et al* (2011) referem que a atitude para reclamar difere das pessoas mais velhas para as mais jovens, demonstrando que esta é uma variável que tem uma dependência com as reclamações. Contudo, também existem estudos contraditórios, pois alguns autores afirmam que os Clientes quanto mais jovens maior é a propensão para reclamar, enquanto outros acreditam que os Clientes a partir dos 54 anos criam maiores expectativas em relação aos produtos e têm mais tempo disponível que os jovens, pelo que têm uma maior tendência para reclamar.

- Poder de compra

O rendimento é claramente uma variável que influencia o poder de compra – consumo de maior quantidade e um leque de produtos com preços mais abrangentes. No seu trabalho, os autores acreditam também que este pode estar relacionado com a atitude para reclamar, assumindo que as pessoas ao terem mais produtos e poder de comparação, têm mais motivos para reclamar. Por sua vez, tal como com as variáveis supracitadas, esta é uma afirmação que gera opiniões contraditórias. Phau and Baird (2008) indica que não existe uma relação significativa entre os diferentes níveis de rendimento, mas no âmbito das diferentes categorias perante a insatisfação, pessoas com maior rendimento reclamam mais facilmente.

---

<sup>1</sup> Decreto-Lei nº74 de 2017

- Educação

A educação é também considerada uma variável importante neste âmbito, pois afeta a forma como os Clientes interpretam e processam o ato de reclamar. De acordo com Liefeld *et al* (1975), quanto maior o nível de educação, maior é a competência para elaborar reclamações, perceber injustiças e as razões para se reclamar. Por sua vez, Phau and Baird (2008), também afirmam que pessoas com um menor nível de educação, reclamam menos. Deste modo, a maioria dos estudos, ao contrário das outras variáveis em que existe discordância, afirmam haver uma relação entre o nível de educação e as reclamações.

Uma vez que existem diversos fatores relacionados com o que leva um Cliente a reclamar, conclui-se que nem todos os Clientes optam por esse caminho. Hirschman (1970), mencionado em (Boote, 1998; Crie, 2003), refere que existem três tipos de comportamento que podem ser adotados pelos Clientes quando estão insatisfeitos: i) Saída (*Exit*); ii) Reclamação (*Voice*); iii) Lealdade (*Loyalty*).

Este último comportamento corresponde aos Clientes que optam por não fazer nada e mantêm-se na empresa por não se sentirem adequadamente qualificados para expressar uma reclamação ou por considerarem que têm ausência do conhecimento necessário para avaliar a qualidade do serviço ou, alternativamente, por sentirem que uma reclamação é uma perda de tempo (Battaglia *et al*, 2010). Através destes as empresas não conseguem retirar qualquer tipo de informação, pelo que os Clientes que reclamam são os únicos que apresentam um comportamento construtivo para a entidade.

Para além destes três tipos, existem outros dois comportamentos relacionados com a reclamação que podem ser adotados: o passa-a-palavra (já apresentado anteriormente), e o recurso a entidades externas (*third-party action*). Enquanto que a saída apresenta uma perda para a organização, um passa-a-palavra negativo pode causar um impacto muito pior, uma vez que pode influenciar muitos outros consumidores. Por outro lado, o recurso a entidades externas, como advogados por exemplo, requer uma maior atenção da parte da empresa e, como resultado, esta toma conhecimento da natureza do problema. Contudo, este comportamento embora apresente vantagens para empresa, não garante que o Cliente não a abandone, devido aos gastos necessários para recorrer a ajudas externas (Davidow *et al*, 1997).

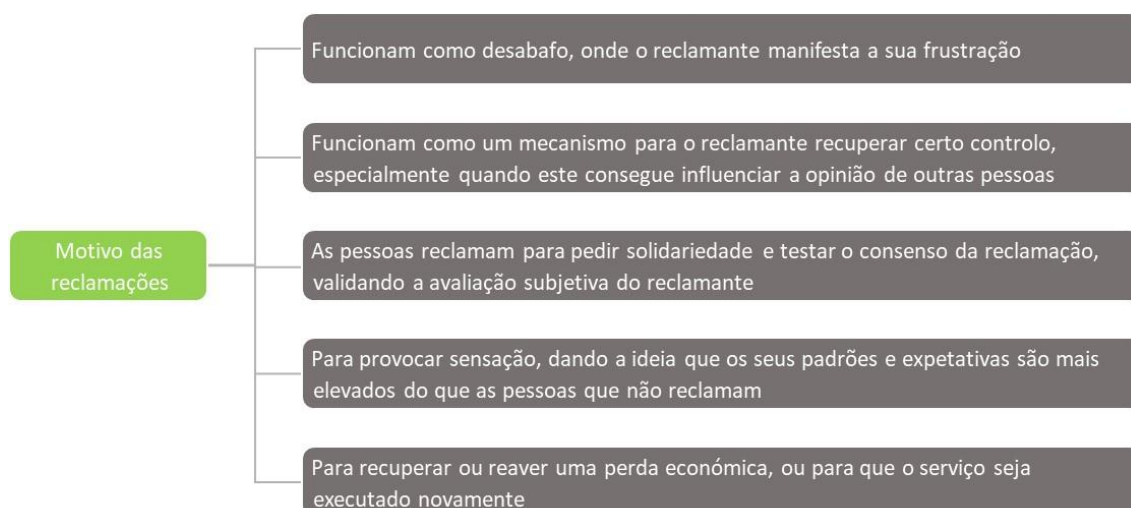
Deste modo, reforça-se a importância de as reclamações serem feitas diretamente à organização, tal como já foi mencionado.

Outros estudos que têm sido feitos estão relacionados com os diversos fatores que podem levar um consumidor a reclamar, de modo a compreenderem o seu comportamento. Day (1984), por exemplo, desenvolveu um modelo relativo à decisão de reclamar (Figura 1.2), onde são analisadas diversas variáveis relacionadas com os custos-benefícios juntamente com variáveis relacionadas com a personalidade do reclamante antes da tomada da decisão.



**Figura 1.2** – Modelo do processo de decisão para reclamar  
Fonte: Adaptado de Day (1984)

Associadas a estas variáveis encontram-se os motivos que o podem levar a reclamar. Na figura 1.3 encontram-se sintetizados os principais motivos.



**Figura 1.3** – Motivos das reclamações  
Fonte: Adaptado de Battaglia *et al* (2010)

Estes motivos são impulsionados pela insatisfação do Cliente. Contudo, estudos como Kowalski (1996), mencionados em Boote (1998), indicam que nem sempre uma reclamação está associada à insatisfação, pois, por vezes, os Clientes podem requerer apenas alguma recompensa da parte dos fornecedores.

Para a elaboração de uma reclamação é exigido tempo e esforço da parte do Cliente, pelo que este espera ser ajudado rapidamente e ser recompensado pela falha ocorrida (Battaglia *et al*, 2010). Estes são alguns dos aspetos que reforçam a importância de se estabelecer um bom processo de gestão de reclamações.

### 1.3.1 Gestão das Reclamações

Segundo Johnston *et al* (2002), uma boa gestão de reclamações e recuperação de Clientes, isto é, lidar com estes após uma falha num serviço, deveria ser o objetivo da estratégia relacionada com a

satisfação dos Clientes de uma organização. Estes autores verificaram também que uma reclamação, embora traga custos para as empresas, acaba por recompensar através dos potenciais benefícios obtidos através da melhoria da prestação dos serviços e também através da retenção de Clientes.

Deste modo, a gestão de reclamações pode ser vista como um investimento, um método de avaliação do desempenho das empresas em relação à satisfação dos seus Clientes com os serviços prestados. Porém, Johnston e Michel (2008), mencionados em Battaglia *et al* (2010), observam que poucas organizações adotam políticas eficazes que permitem entender as reclamações dos seus Clientes e, assim, identificar a origem dos problemas.

Numa perspetiva de processo, o tratamento de uma reclamação pode ser visto como uma sequência de eventos que se inicia com a comunicação do Cliente, gerando assim um conjunto de interações através do qual termina com uma decisão (Tax *et al*, 1998).

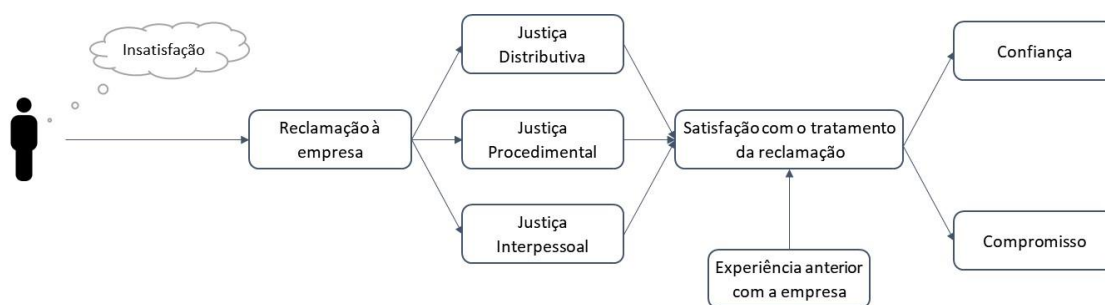
Porém, numa situação de dupla insatisfação, é possível gerarem-se novas reclamações, caso os Clientes não se sintam satisfeitos com a experiência e tratamento da reclamação.

Tax *et al* (1998) analisaram todo este processo e qual o impacto que pode ter uma má gestão de reclamações num Cliente insatisfeito, através da teoria da justiça, concluindo que os Clientes avaliam este processo com base em: resultados, procedimento utilizado até chegar ao resultado e ainda ao tratamento interpessoal durante a sua execução.

Segundo estes autores, o conceito de justiça é constituído por três elementos:

- justiça distributiva – resultado da decisão com foco nos custos-recompensas por parte da organização em resposta ao problema gerado;
- justiça procedimental – forma como o problema é solucionado, sendo composta por: controlo de processo e decisão, tempo/rapidez da resposta, acessibilidade e flexibilidade;
- justiça interpessoal – comportamentos interpessoais com o reclamante por parte da organização e dos seus funcionários.

O processo de gestão de reclamações, criado com base nesta teoria (figura 1.4), permite concluir que um processo que transmite justiça ao Cliente aliado às experiências positivas anteriores com a empresa, resultam num sentimento de confiança e compromisso da parte do Cliente (Tax *et al*, 1998).



**Figura 1.4** – Processo de gestão da reclamação com base na justiça  
Fonte: Adaptado de Tax *et al* (1998)

## 2. O Negócio

### 2.1 O Setor Elétrico em Portugal

A energia elétrica é uma das mais importantes descobertas feitas pelo ser humano, gerando um grande contributo para a evolução da humanidade e do mundo.

Em Portugal, a energia, tal como é conhecida, surgiu há pouco mais de um século e tornou-se fundamental para o quotidiano. No final do século XIX, começaram a surgir as primeiras empresas de produção e distribuição de energia elétrica: as Companhias Reunidas de Gás e Eletricidade (1891) e algumas centrais (Rollo, 2015).

Em 1969 é constituída a Companhia Portuguesa de Eletricidade (CPE), fruto da fusão das concessionárias da produção e transporte da rede elétrica existentes até à data.

Anos mais tarde, após o 25 de abril, o país atravessava um desequilíbrio na eletrificação do território continental, pois esta concentrava-se nos grandes centros urbanos, estando o resto do país sem acesso à eletricidade, pelo que este foi o principal desafio encontrado pela EDP – Eletricidade de Portugal (1976), empresa que surgiu da fusão de 13 das 14 empresas de produção, transporte e distribuição que haviam sido nacionalizadas em 1975 (Antena 1, 2015; EDP, 2016), tornando-se, assim, detentora do monopólio natural deste setor.

Em 1984, as redes de distribuição de Alta e Média Tensão já cobriam todo o território, enquanto que a rede de Baixa Tensão atingia 92%.

Com a entrada de Portugal na CEE (Comunidade Económica Europeia), a presença de um monopólio neste setor revelou-se um obstáculo para a evolução do mesmo, pelo que a União Europeia determinou que o mercado da energia deveria ser livre, oferecendo ao consumidor a possibilidade de escolher o seu fornecedor, levando assim as áreas da produção e comercialização para o mercado concorrencial (EDP, 2016).

Em 1995, ocorreu a separação das operações de produção, distribuição e transporte de energia, promovendo, uma vez mais, a concorrência nos segmentos competitivos da cadeia de valor do setor elétrico (Figura 2.1). Esta cadeia é constituída pelas 4 fases (e respetivos intervenientes) que a eletricidade atravessa até chegar à casa do consumidor.



**Figura 2.1** – Descrição da cadeia de valor do setor elétrico  
Fonte: Elaboração própria

Na atividade de produção, a eletricidade é produzida através de diferentes tecnologias e fontes primárias de energia. Em Portugal, as fontes mais utilizadas são o carvão, o gás natural, o vento (energia eólica) e a água (energia hídrica).

Depois de produzida, a eletricidade é transportada até às subestações de distribuição ou até a algumas instalações que necessitam de eletricidade em Muito Alta Tensão (por exemplo, ferroviárias). Em Portugal continental, esta atividade é da responsabilidade da Rede Elétrica Nacional (REN) que atua isoladamente, fazendo o transporte da energia de Muito Alta Tensão entre os produtores e os distribuidores. Em 2000, o Governo Português decidiu garantir o controlo público da REN, separando-a assim do Grupo EDP, uma vez que num contexto de mercado concorrencial tinha que transportar energia de outros produtores para além da EDP Produção, evitando-se assim conflitos de interesse, garantido a confiança dos consumidores e produtores.

Nas subestações de distribuição e nos postos de transformação, dá-se a redução da tensão da energia elétrica para que esta possa ser distribuída até aos consumidores finais. A EDP Distribuição é a principal responsável pela distribuição desta eletricidade, existindo também algumas concessões de energia em baixa tensão com esta função.

Por fim, a atividade de comercialização assegura a venda desta energia ao consumidor, sendo este livre de escolher o seu comercializador. Atualmente, existem dois tipos de comercializadores: os comercializadores livres (mercado liberalizado) e os comercializadores de último recurso (mercado regulado).

Com o objetivo de dar mais poder de decisão e intervenção ao Cliente, surge também a ERSE (Entidade Reguladora dos Serviços Energéticos) com o papel de proteção dos interesses do mesmo, sendo responsável pela regulação deste setor, tendo como objetivos promover a eficiência das empresas, a qualidade do serviço, a concorrência entre os agentes intervenientes, entre outros (ERSE, 2017a).

No mercado regulado, os preços de venda de energia para os consumidores finais são fixados pela ERSE. Enquanto que no mercado livre são os comercializadores a decidir os preços e as condições comerciais, respeitando as regras e os regulamentos estipulados.

Tratando-se este de um setor com um papel fundamental tanto a nível social como económico é classificado também como um Serviço de Interesse Económico Geral (SIEG). Estes serviços são classificados como sendo de interesse geral pelas autoridades públicas dos Estados-Membros da União Europeia, estando sujeitos a obrigações específicas de serviço público (ver Comissão Europeia, 2013).

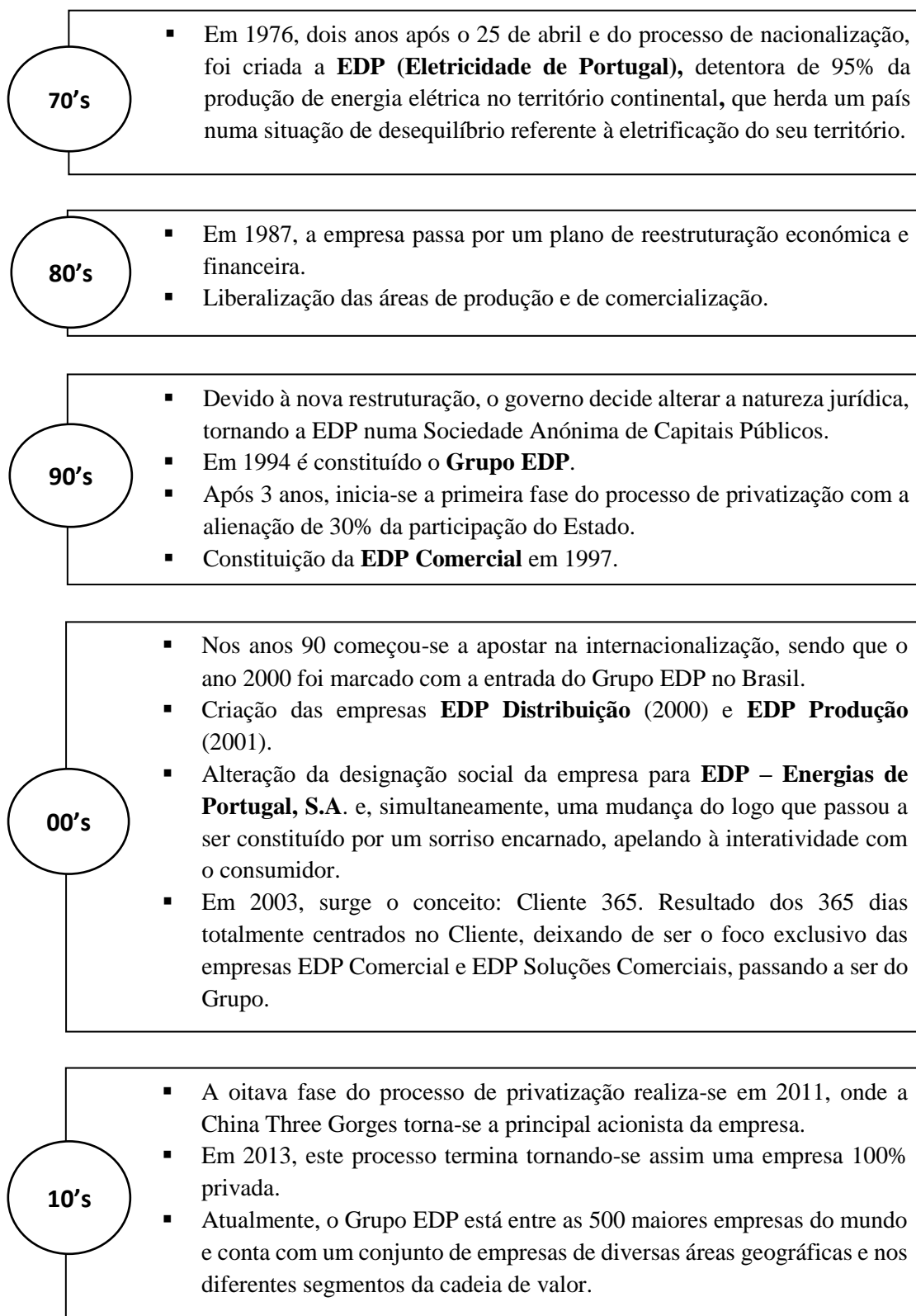
## **2.2 O Grupo EDP**

### **2.2.1 A Visão**

O Grupo EDP tem como visão ser “uma empresa global de energia, líder em criação de valor, inovação e sustentabilidade”, sendo sustentada pelos seguintes compromissos (EDP, 2017):

- i) Com as pessoas – Valorizando o trabalho em equipa e promovendo o desenvolvimento de competências e mérito;
- ii) Com os Clientes – Colocando-se no lugar do Cliente na tomada de decisões, ouvindo-os e respondendo-lhes de forma simples e transparente, antecipando as suas necessidades;
- iii) Com a sustentabilidade – Assumindo as responsabilidades sociais e ambientais, simultaneamente que promovem ativamente a eficiência energética;
- iv) Com os resultados – Cumprindo os compromissos que assumem perante os acionistas e exigindo a excelência em tudo o que fazem.

### 2.2.2 Evolução Histórica



**Figura 2.2** – História do Grupo EDP

Fonte: Elaboração própria



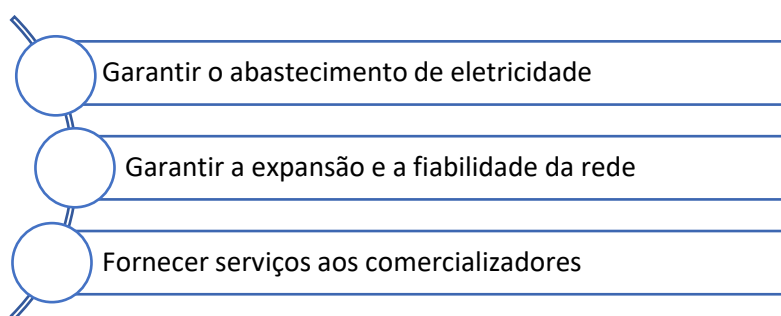
## 2.3 EDP Distribuição

A EDP Distribuição nasceu em fevereiro de 2000, fusão das quatro empresas que constituíam o universo da distribuição da energia elétrica do Grupo EDP.

Devido às novas condições regulatórias da liberalização do setor, a EDP Distribuição também teve de se adaptar, passando a ser constituída por três áreas distintas de negócio: Gestão de Redes, Serviço de Redes e Comercial (EDP, 2016).

Atualmente, detém 99% da rede de distribuição de energia elétrica em Portugal continental, sendo responsável por mais de 20% dos resultados líquidos do Grupo, contando com mais de 6 milhões de Clientes.

A EDP Distribuição, na cadeia de valor, tem o papel de fazer a conexão entre a REN e os consumidores finais que têm uma relação contratual com os comercializadores. É-lhe atribuída o papel de Operador de Redes de Distribuição (ORD), contemplando assim as seguintes atividades:



**Figura 2.3** – Tipos de atividades que constituem o negócio de distribuição de energia elétrica  
Fonte: Elaboração própria

Para além destas obrigações, a EDP Distribuição também reconhece o seu papel como agente facilitador do desenvolvimento do setor elétrico e da relação dos Clientes com o consumo de energia (EDP Distribuição, 2017a).

E, por pertencer a um grupo empresarial que detém tanto empresas de produção como comercialização, encontra-se ainda sujeita a obrigações complementares relacionadas com a independência (não discriminação e separação de atividades), transparência e confidencialidade (EDP Distribuição, 2017b).

A EDP Distribuição está organizada em direções de âmbito nacional e regional, sendo que as Direções de Rede e Concessões<sup>2</sup> (DRC) se distribuem conforme apresentado na figura 2.4.

E, por sua vez, cada DRC é constituída por diversas Áreas de Rede e Clientes<sup>3</sup> (ARC) e gabinetes.

Por exemplo, a DRC Lisboa (DRCL) é constituída pelas ARC Lisboa, ARC Setúbal e ARC Loures.

<sup>2</sup> Antigas Direções de Redes e Clientes

<sup>3</sup> Antigas Áreas Operacionais



**Figura 2.4** – Direções de Redes e Concessões  
Fonte: EDP (2017c)

## 3. Metodologias

### 3.1 Introdução

Atualmente, para qualquer área de negócio há a necessidade de utilizar diversos métodos rigorosos para a tomada de decisões para diferentes situações ou problemas que possam surgir. A estatística tem então sido fundamental para ajudar as empresas nos tratamentos dos seus dados, ajudando a tomar as decisões mais acertadas. Esta encarrega-se por estudar a variabilidade apresentada pelos dados e, a partir destes, permite extrair algumas conclusões com um certo nível de confiança.

Um dos ramos desta área é a inferência estatística, que permite tirar conclusões relacionadas com uma população com base numa amostra extraída dessa mesma população, recorrendo a testes estatísticos paramétricos ou não paramétricos, técnicas de estimação e intervalos de confiança, entre outros.

Segundo Diez *et al* (2014) é interessante tratar a estatística como um processo de investigação, onde devem ser adotados os seguintes passos:

1. Identificar a questão ou problema;
2. Selecionar os dados relevantes para o problema;
3. Analisá-los;
4. Formular uma conclusão.

Seguindo este raciocínio, todo este processo inicia-se com a identificação da questão/problema e seleção dos dados necessários. Posteriormente, dedica-se à análise dos mesmos.

Para a elaboração de uma boa análise, ter os dados necessários não é suficiente. Nos dias de hoje e, principalmente através da evolução das tecnologias, é cada vez maior o leque de testes estatísticos disponíveis e a sua facilidade de aplicação, pelo que se torna cada vez mais fundamental conhecer os tratamentos mais adequados e as respetivas limitações, consoante o tipo de dados em estudo (designados também por variáveis) e a informação que se pretende retirar.

### Tipo de Dados

Relativamente ao tipo de dados, numa primeira fase estes podem ser classificados como numéricos (quantitativos) ou categóricos (qualitativos).

Na realização de um estudo estatístico, os dados nem sempre são apresentados num formato numérico, impossibilitando assim qualquer quantificação. A esses dados atribui-se o nome de dados categóricos que, tal como o nome indica, são dados agrupados por um número finito de categorias, como por exemplo, “Norte” e “Sul” que correspondem a duas categorias. Estes são essencialmente utilizados para medir atitudes, opiniões e/ou respostas, tendo sido desenvolvidos para estimular estudos relacionados com as ciências sociais ou biomédicas (Agresti, 2002).

Para além desta distinção, as variáveis podem ainda ser agrupadas consoante as suas características.

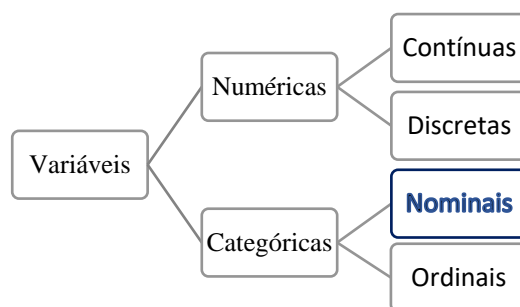
As variáveis categóricas ordinais são aquelas que apresentam uma ordem natural, sem qualquer indicação, como por exemplo: “nenhum”, “algum” e “muito”. Por outro lado, as variáveis categóricas nominais são aquelas que não apresentam esta característica, sendo indiferente a ordem em que estas surgem.

Por sua vez, as variáveis numéricas podem ser classificadas como contínuas ou discretas, de acordo com o número de valores que podem assumir. As variáveis discretas só consideram um conjunto numerável de valores, sendo geralmente utilizadas para resultados de contagens, como por exemplo, o

número de reclamações em 2017, ao passo que as variáveis contínuas são medidas através de uma escala contínua, podendo assumir todos os valores reais, como por exemplo, a altura de um Cliente.

Este trabalho irá concentrar-se essencialmente nas variáveis categóricas nominais, onde uma das variáveis em estudo está relacionada com a presença ou não de reclamações face ao tema “visitas não realizadas”. As respostas possíveis serão “sim” ou “não”, associando os valores “1” ou “0” a cada categoria, respetivamente. Às variáveis nestas condições, onde existem apenas duas categorias possíveis é-lhes atribuído o nome de variáveis binárias.

Na figura abaixo segue um resumo dos tipos de variáveis que se podem encontrar em qualquer teste estatístico:



**Figura 3.1** – Tipos de variáveis  
Fonte: Elaboração própria

As variáveis, para além de serem classificadas com base no seu tipo, podem ser também classificadas com base na relação entre si, relação esta denominada por associação. Uma variável diz-se dependente (ou resposta) quando o seu valor depende do comportamento de outras variáveis - variáveis independentes (ou explicativas).

### 3.2 Inferência Estatística

Muitas questões relacionadas com dados categóricos podem ser respondidas envolvendo inferência estatística. Para isto, podem-se recorrer a testes de hipóteses e estimação de parâmetros. Neste trabalho são utilizados alguns, tais como estimação de máxima verosimilhança ou testes do qui-quadrado, de modo a perceber o tipo de relação entre as diversas variáveis em estudo.

#### Testes de Hipóteses

Em muitas situações práticas requer-se que se tomem decisões relacionadas com os parâmetros (ou outros aspetos) da população em função de valores observados numa determinada amostra. Para isso recorrem-se aos testes de hipóteses, que têm como objetivo conduzir o investigador a uma decisão com base nessa amostra recolhida.

Estes testes são constituídos por duas hipóteses: a hipótese nula,  $H_0$ , e a hipótese alternativa,  $H_1$ . A hipótese alternativa está associada à hipótese em investigação, isto é, o que se pretende verificar no contexto em análise, enquanto que a hipótese nula corresponde à hipótese complementar a  $H_1$ . Estas hipóteses podem ser constituídas por diferenças ou igualdades, resultando em testes unilaterais (quando a hipótese alternativa apresenta os sinais de maior ou menor) ou bilaterais (nos casos em que a hipótese alternativa possui um sinal de diferente).

A rejeição de  $H_0$  ocorre sempre que o valor da estatística de teste, obtida através dos dados da amostra, pertence à região crítica, concluindo-se que existe evidência contra esta hipótese.

A região crítica é construída com base no valor crítico ou nível de confiança. Este está associado ao tipo de erros que se podem cometer ao tomar uma decisão, uma vez que o teste é baseado em amostras. Por exemplo, sempre que se afirma algo com 95% de confiança, significa que existe 5% de probabilidade de rejeitar a hipótese nula sendo esta verdadeira - erro tipo I. Correspondendo a um nível de significância de  $\alpha=0,05$ ,  $\alpha$  representa a probabilidade de ocorrer este tipo de erro.

Porém, tal como existe a probabilidade de ocorrer este tipo de erro (isto é, obter um “falso positivo”), também existe a probabilidade de não se rejeitar a hipótese nula quando esta não é verdadeira - erro tipo II (isto é, obter um “falso negativo”), representado por  $\beta$ . Assim, surge também o conceito de potência, que corresponde à probabilidade de rejeitar  $H_0$  quando esta é falsa. Esta potência é dada por  $1-\beta$ . Deste modo, em qualquer estudo estatístico, é desejável um nível de significância baixo, assim como um teste potente.

Os custos de se obterem falsos positivos ou falso negativos variam consoante o experimento em causa e, deste modo, torna-se importante perceber primeiramente quais as consequências de se cometerem os dois tipos de erro, decidindo qual o valor para o nível de significância mais adequado. Contudo, em termos práticos esta não é uma decisão fácil de se tomar, então atribui-se, geralmente, os valores de 0,01, 0,05 e 0,10 para  $\alpha$ , sendo 0,05 o preferido de muitos autores (Agresti, 2002).

### ***P-value***

O *p-value* (em português, valor-*p*) está associado ao nível de significância de um teste de hipóteses. Este corresponde à probabilidade de a estatística de teste ter o valor máximo em relação ao valor observado, quando a hipótese nula é verdadeira.

Quando o *p-value* é menor que o nível de significância proposto, o valor obtido para a estatística de teste encontra-se na região crítica, ao nível  $\alpha$ , e portanto, rejeita-se a hipótese nula. Outra interpretação para este valor é corresponder ao menor valor usual do nível de significância para o qual se rejeita a hipótese nula.

### **Testes não paramétricos**

Dentro dos testes de hipóteses, estes podem-se distinguir entre dois tipos: testes paramétricos e não paramétricos.

Os testes paramétricos incidem sobre um ou mais parâmetros das populações, exigindo a sua normalidade, especialmente quando a(s) amostra(s) têm uma dimensão reduzida. No entanto, estes pressupostos nem sempre são verificados. Assim, surgem os testes não paramétricos com o sentido de colmatar estes casos, podendo ser aplicados em amostras de grandes ou pequenas dimensões, ou em dados de natureza qualitativa.

Na tabela 3.1 é apresentado um breve resumo das principais vantagens e desvantagens destes testes face aos testes paramétricos.

**Tabela 3.1** – Testes não paramétricos: vantagens e desvantagens

| <b>Testes não paramétricos</b>   |   |
|--|---|
| <u><b>Vantagens</b></u>  | <u><b>Desvantagens</b></u>  |
| <ul style="list-style-type: none"> <li>- Podem ser aplicados a dados não numéricos;</li> <li>- Poucos pressupostos;</li> <li>- Simplicidade na aplicação.</li> </ul> | <ul style="list-style-type: none"> <li>- Pode verificar-se perda de informação ao comparar com os testes paramétricos;</li> <li>- Apresentam uma menor potência;</li> <li>- Para amostras de grandes dimensões são tendenciosos e trabalhosos.</li> </ul> |

Os primeiros testes desenvolvidos em estatística não paramétrica foram atribuídos a John Arbuthnot (1710) por Lehmann, mencionado em Contador *et al* (2016), pela apresentação do teste dos sinais. Pearson, em 1900, deu um grande passo relativamente a este tema, inserindo também uma aplicação para as variáveis nominais e categóricas através do teste do qui-quadrado. Mais tarde, surgiram variadas abordagens para o caso das variáveis ordinais, graças a Wilcoxon (1945), Mann & Whitney (1947), Kruskal & Wallis (1952), entre outros. Desde aí que a estatística não paramétrica tem vindo a representar um ramo extremamente importante na estatística até aos dias de hoje (Contador *et al*, 2016).

Neste trabalho, as variáveis em estudo serão sempre de natureza qualitativa, pelo que o seu foco serão os testes não paramétricos, destacando-se os testes para proporções e os testes do qui-quadrado.

## **Distribuição Binomial**

A distribuição binomial é o modelo probabilístico adequado para os casos em que se considera um conjunto de provas repetidas de Bernoulli, isto é, experiências sucessivas aleatórias e independentes onde se verifica a presença ou não de determinada característica com probabilidades  $\pi$  e  $1-\pi$ , respetivamente.

Seja  $Y$  a variável aleatória que designa o número de sucessos das  $n$  provas independentes e identicamente distribuídas (i.i.d.), isto é,  $Y = \sum_{i=1}^n X_i$ , onde  $\{X_i\}$  representam as variáveis resposta das extrações de Bernoulli. Por isso,  $Y$  segue uma distribuição binomial com parâmetros  $(n, \pi)$ , cuja função de probabilidades é dada por:

$$P(y) = \binom{n}{y} \pi^y (1-\pi)^{n-y}, y = 0, 1, \dots, n, \text{ onde } \binom{n}{y} = \frac{n!}{y!(n-y)!}. \quad (3.1)$$

O valor médio e a variância desta distribuição são  $E[Y] = n\pi$  e  $var[Y] = n\pi(1-\pi)$ , respetivamente.

Na presença de variáveis categóricas nem sempre são disponibilizados todos os parâmetros necessários, pelo que é preciso recorrer à estimação das proporções dos mesmos. Para esta estimação usa-se a inferência estatística, utilizando métodos tais como o cálculo do estimador de máxima verosimilhança.

No caso da distribuição binomial, com  $Y$  sucessos numa amostra de dimensão  $n$ , o estimador de máxima verosimilhança de  $\pi$  é dado por  $\hat{\pi} = \frac{Y}{n}$ , como se demonstra seguidamente, que corresponde à proporção de indivíduos com a presença da característica de interesse na amostra.

*Demonstração:* Sejam  $(X_1, \dots, X_n)$   $n$  variáveis i.i.d com distribuição Bernoulli ( $\pi$ ); Considere-se  $Y = \sum_{i=1}^n X_i$  uma variável que segue uma distribuição binomial  $(n, \pi)$ , como definida em (3.1).

A função log-verossimilhança da amostra  $\underline{y} = (y_1, \dots, y_k)$  é dada por:

$$\log L(\pi | \underline{y}) = \sum_{i=1}^k y_i \log(\pi) + \sum_{i=1}^k (1 - y_i) \log(1 - \pi). \quad (3.2)$$

Derivando (3.2) em ordem a  $\pi$  e igualando a zero, obtém-se:

$$\hat{\pi} = \frac{\sum_{i=1}^k y_i}{nk}. \quad (3.3)$$

Para  $k=1$ , conclui-se que o estimador de máxima verossimilhança para o parâmetro  $\pi$  é:

$$\hat{\pi} = \frac{Y}{n}. \quad (3.4)$$

□

Os estimadores de máxima verossimilhança têm bastante utilidade para os mais diversos estudos estatísticos, uma vez que para além de estimarem o valor dos parâmetros desconhecidos, estes também possuem propriedades muito importantes quando respeitam certas condições.

Uma das condições mais comum que se exige de um estimador é que este seja centrado (ou não enviesado), isto é, que o seu valor esperado coincida com o parâmetro a estimar.

Tendo em conta o supracitado, facilmente se demonstra que  $\hat{\pi}$  é um estimador centrado para  $\pi$ . Como  $Y$  tem uma distribuição binomial, vem que  $E[Y] = n\pi$  e  $Var[Y] = n\pi(1 - \pi)$ . Então:

$$E[\hat{\pi}] = E\left[\frac{Y}{n}\right] = \frac{1}{n} E[Y] = \pi. \quad (3.5)$$

Outra condição que também se pode exigir está relacionada com a consistência do estimador, isto é, que este convirja em probabilidade para o valor do parâmetro.

Mais uma vez, facilmente se verifica que  $\hat{\pi}$  é um estimador consistente. Como  $\hat{\pi}$  é dado por  $\frac{Y}{n}$  e este corresponde à média amostral de  $Y$ , então pela Lei Fraca dos Grandes Números, é imediato que este é um estimador consistente para o parâmetro, pois a média amostral de  $Y$  converge em probabilidade para o parâmetro  $\pi$ . (Note-se que este resultado apenas é válido para valores razoavelmente grandes de  $n$ ).

## Intervalo de Confiança

Uma alternativa a este método, em que se encontra um único valor como estimativa de um determinado parâmetro, é através da construção de um intervalo ao qual o verdadeiro valor do parâmetro pertence, com uma certa probabilidade. Em termos práticos, isto quer dizer que em sucessivas amostragens, o referido intervalo contém o parâmetro numa percentagem de vezes que se pretende elevada. A este método atribui-se o nome de intervalos de confiança para o parâmetro em causa que, na prática, é uma alternativa que disponibiliza mais informação (Agresti, 2002).

Uma das formas mais utilizada para determinar estes intervalos de confiança é através do método da variável fulcral.

Considere-se a amostra aleatória  $\underline{X} = (X_1, \dots, X_n)$  proveniente de uma população com distribuição  $F(x; \pi)$ . Assumindo que  $S(\pi; \underline{X})$  é uma variável aleatória que representa uma função da amostra e do

parâmetro desconhecido  $\pi$ , então diz-se que esta é uma variável fulcral se a sua distribuição não depender do parâmetro  $\pi$  e este ser o único valor desconhecido.

Deste modo, podem ser determinados os valores  $a$  e  $b$ , tais que:

$$P[a < S(\pi; \bar{X}) < b] = 1 - \alpha. \quad (3.6)$$

Com o intuito de construir um intervalo com  $(1-\alpha)100\%$  de confiança para o valor médio ( $\mu$ ) de uma população normal com variância ( $\sigma^2$ ) conhecida, é possível utilizar-se a variável fulcral

$$\sqrt{n} \frac{\bar{X} - \mu}{\sigma}, \quad (3.7)$$

que depende apenas da quantidade desconhecida  $\mu$  e que a distribuição de probabilidade tem valor médio nulo e variância unitária.

Seja  $q_{1-\alpha/2}$  o quantil de ordem  $1-\alpha/2$  desta distribuição, tal que:

$$\Phi(q_{1-\alpha/2}) = 1 - \frac{\alpha}{2}. \quad (3.8)$$

Então, substituindo  $a$  e  $b$  pelos respectivos valores, obtém-se o intervalo de confiança para  $\mu$  através da fórmula:

$$P[-q_{1-\alpha/2} < \sqrt{n} \frac{\bar{X} - \mu}{\sigma} < q_{1-\alpha/2}] = 1 - \alpha \Leftrightarrow P[\bar{X} - \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2}] = 1 - \alpha. \quad (3.9)$$

É importante realçar o facto que quanto maior for o grau de confiança escolhido, maior será a amplitude média do intervalo obtido. E, por sua vez, quanto maior a dimensão da amostra, menor será a amplitude média deste, ou seja, maior será a precisão do estimador.

Nos próximos subcapítulos, irá compreender-se a utilidade de se perceber como funcionam os intervalos de confiança para populações normais, uma vez que pelo Teorema do Limite Central estes também são válidos para populações não normais.

### 3.2.1 Teste para uma Proporção

Seja  $\pi$  a proporção de indivíduos com a característica de interesse. Suponha-se que se pretende testar se este valor é inferior a uma dada proporção, representada por  $\pi_0$ , isto é,

$$H_0: \pi \geq \pi_0 \text{ vs. } H_1: \pi < \pi_0. \quad (3.10)$$

Tendo por base a variável aleatória  $Y$  com distribuição considerada em (3.1) e para um valor de  $n$  razoavelmente grande, pode-se concluir que o estimador  $\hat{\pi}$  do parâmetro  $\pi$  aproxima-se de uma distribuição normal, através do Teorema do Limite Central, com valor médio e desvio-padrão dados, respetivamente, por:

$$E[\hat{\pi}] = \pi, \sigma[\hat{\pi}] = \sqrt{\frac{\pi(1-\pi)}{n}}. \quad (3.11)$$



Assim, é válido o seguinte resultado:

$$Z = \frac{\hat{\pi} - \pi}{\sqrt{\pi \frac{1-\pi}{n}}} \xrightarrow[n \rightarrow \infty]{d} N(0,1). \quad (3.12)$$

Então, sob  $H_0$ , o valor observado para  $\hat{\pi}$ , denotado por  $Z_0$ , é dado por:

$$Z_0 = \frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0 \frac{1-\pi_0}{n}}}. \quad (3.13)$$

Testar a hipótese nula  $H_0: \pi \geq \pi_0$  contra a hipótese alternativa  $H_1: \pi < \pi_0$  corresponde à probabilidade de se observar um valor de  $\hat{\pi}$  tão ou mais pequeno que o valor realmente observado, assumindo a hipótese nula. Por isso, o  $p$ -value para estas hipóteses corresponde a  $P(Z \leq Z_0) \simeq \Phi(Z_0)$ , onde se rejeita  $H_0$  quando este valor é menor ou igual ao  $\alpha$  estabelecido.

Note-se que se fosse um teste bilateral, isto é, caso se pretendesse testar  $H_0: \pi = \pi_0$  contra  $H_1: \pi \neq \pi_0$ , o  $p$ -value seria calculado através de  $P[|Z| \geq |Z_0|] \simeq 2 \times (1 - \Phi(|Z_0|))$ , onde também se rejeita  $H_0$  para valores menores ou iguais ao  $\alpha$  estabelecido.

## Intervalo de Confiança

Anteriormente, a título de exemplo, verificou-se como se constrói um intervalo de confiança para amostras de populações com distribuição normal. Embora existam testes específicos baseados em distribuições binomiais, estes apenas funcionam quando os intervalos de confiança não são simétricos em torno da proporção, especialmente quando esta está próxima de zero ou um, pelo que assumir a normalidade das proporções é uma alternativa mais simples de aplicar quando se trabalha com variáveis nominais, garantido intervalos de confiança simétricos, embora não funcione para proporções próximas de zero ou um (McDonald, 2014). Contudo, dado que se trata de uma amostra de grandes dimensões, este método é considerado o mais apropriado.

Quando os estimadores de máxima verosimilhança são centrados (ou assintoticamente centrados) podem ser utilizados como variáveis fulcrais, seguindo assintoticamente uma distribuição normal. Esta propriedade é bastante útil, uma vez que nem sempre é fácil encontrar uma variável fulcral cuja distribuição seja conhecida.

Dado que  $\hat{\pi}$  é um estimador centrado para o parâmetro  $\pi$ , então a variável  $Z$  (apresentada em (3.12)) pode ser utilizada como variável fulcral, permitindo assim a construção de um intervalo de confiança para  $\pi$ , com grau de confiança  $1-\alpha$ :

$$\hat{\pi} \pm q_{1-\alpha/2} \sqrt{\frac{\pi(1-\pi)}{n}}. \quad (3.14)$$

Contudo, este intervalo depende de parâmetros desconhecidos. Em condições bastante gerais, a distribuição de  $Z$  não se altera significativamente na substituição do valor do parâmetro por um estimador seu quando este é um estimador consistente.

Neste caso, pela Lei Fraca dos Grandes Números sabe-se que  $\hat{\pi}$  converge, em probabilidade, para  $\pi$  e, sendo  $\sqrt{\pi(1-\pi)}$  uma função contínua, então  $\sqrt{\hat{\pi}(1-\hat{\pi})}$  também converge, em probabilidade, para  $\sqrt{\pi(1-\pi)}$ . Consequentemente, pelo teorema de Slutsky, substituindo  $\pi$  por  $\hat{\pi}$  em (3.14), tem-se:

$$\hat{\pi} \pm q_{1-\alpha/2} \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}. \quad (3.15)$$

### 3.2.2 Teste para a Diferença de Proporções

Sejam  $\pi_1$  e  $\pi_2$  as probabilidades de se selecionar um indivíduo com certa característica em duas populações independentes. Suponha-se que se pretende comparar as duas proporções nas diferentes amostras com dimensões são  $n_1$  e  $n_2$ , respetivamente.

Então,  $\hat{\pi}_1$  e  $\hat{\pi}_2$  representam as proporções correspondentes, em que  $\hat{\pi}_i$  é a proporção amostral com essa característica na amostra  $i$ , para  $i=1,2$ . Como  $\hat{\pi}_1$  e  $\hat{\pi}_2$  são estimadores não enviesados de  $\pi_1$  e  $\pi_2$ , pode-se concluir que  $\hat{\pi}_1 - \hat{\pi}_2$  é também um estimador não enviesado de  $\pi_1 - \pi_2$ .

Uma vez que se pretende comparar as duas proporções, as hipóteses a testar são:

$$H_0: \pi_1 = \pi_2 \text{ vs. } H_1: \pi_1 \neq \pi_2. \quad (3.16)$$

Sendo as duas amostras independentes, o erro padrão estimado para a diferença das proporções é igual a

$$S.E.(\hat{\pi}_1 - \hat{\pi}_2) = \sqrt{\left( \frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2} \right)}. \quad (3.17)$$

Sob a validade da hipótese nula, é considerada a igualdade das proporções pelo que se pode agregar a informação de ambas as variáveis numa única -  $\hat{\pi}$ , que representa a proporção de sucessos presente nas duas amostras. A esta variável atribui-se o nome de estimador agrupado (= "pooled estimate") e é utilizado nestas situações em que o objetivo se prende em testar a igualdade das proporções.

Substituindo os valores iniciais das proporções na igualdade (3.17) por  $\hat{\pi}$ , obtém-se:

$$S.E.(\hat{\pi}_1 - \hat{\pi}_2) = \sqrt{\hat{\pi}(1-\hat{\pi}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}. \quad (3.18)$$

Então considera-se  $\hat{\pi}$ :

$$\hat{\pi} = \frac{\# \text{ Total de Sucessos nas 2 amostras}}{\# \text{ Total de Observações nas 2 amostras}} = \frac{n_1 \hat{\pi}_1 + n_2 \hat{\pi}_2}{n_1 + n_2}. \quad (3.19)$$

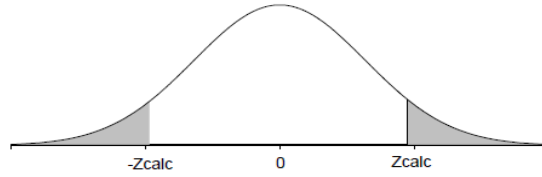
Usando um raciocínio análogo ao aplicado em (3.12), tem-se que:

$$Z^* = \frac{(\hat{\pi}_1 - \hat{\pi}_2) - (\pi_1 - \pi_2)}{S.E.(\hat{\pi}_1 - \hat{\pi}_2)} \xrightarrow[n_2 \rightarrow \infty]{n_1 \rightarrow \infty} N(0,1). \quad (3.20)$$

Pelo que, para se testar as hipóteses  $H_0: \pi_1 = \pi_2$  vs.  $H_1: \pi_1 \neq \pi_2$  e, sob  $H_0$ , isto é, assumindo a igualdade das proporções, o valor observado  $Z_{\text{calc}}$  é dado por:

$$Z_{\text{calc}} = \frac{(\hat{\pi}_1 - \hat{\pi}_2)}{S.E.(\hat{\pi}_1 - \hat{\pi}_2)}. \quad (3.21)$$

Assim, o *p-value* irá corresponder à área abaixo da curva da função densidade normal-padrão tal que:  
 $P[|Z^*| \geq |Z_{\text{calc}}|] \simeq 2 \times (1 - \Phi(|Z_{\text{calc}}|))$ , como se observa na figura 3.2.

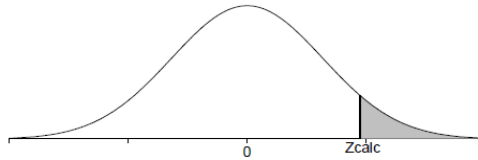


**Figura 3.2** - *P-value* para as hipóteses  $H_0: \pi_1 = \pi_2$  vs.  $H_1: \pi_1 \neq \pi_2$ .  
 Fonte: Berry *et al* (2008)

Contudo, o objetivo do estudo nem sempre passa por testar a diferença das proporções. Quando se pretende analisar se uma proporção é inferior a outra, testam-se as hipóteses:

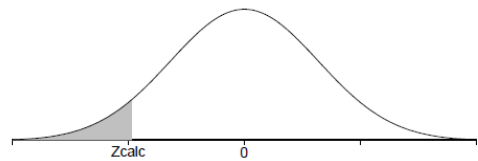
$$H_0: \pi_1 \leq \pi_2 \text{ vs. } H_1: \pi_1 > \pi_2. \quad (3.22)$$

Deste modo, quanto maior for o valor de  $Z_{\text{calc}}$  maior será a evidência contra a hipótese nula e, por sua vez, o *p-value* irá corresponder à área por baixo da curva da função densidade normal-padrão para o lado direito de  $Z_{\text{calc}}$  ( $p\text{-value} = P(Z^* \geq Z_{\text{calc}})$ ) (Figura 3.3).



**Figura 3.3** - *P-value* para as hipóteses  $H_0: \pi_1 \leq \pi_2$  vs  $H_1: \pi_1 > \pi_2$ .  
 Fonte: Berry *et al* (2008)

Caso o pretendido fosse testar a hipótese contrária, isto é, analisar se uma proporção é superior à outra, tanto os sinais das hipóteses como os da probabilidade para se calcular o *p-value* seriam invertidos. Assim, quanto menor o valor de  $Z_{\text{calc}}$ , maior seria a evidência contra a hipótese nula (Figura 3.4).



**Figura 3.4** - *P-value* para as hipóteses  $H_0: \pi_1 \geq \pi_2$  vs  $H_1: \pi_1 < \pi_2$ .  
 Fonte: Berry *et al* (2008)

## Intervalo de Confiança

De uma forma análoga à construção do intervalo de confiança para uma proporção é também possível obter-se um intervalo de confiança para a diferença entre duas proporções.

Considere-se a variável fulcral utilizada em (3.20),  $Z^*$ .

Um intervalo assintótico de grau  $1-\alpha$  é obtido por:

$$\hat{\pi}_1 - \hat{\pi}_2 \pm q_{1-\alpha/2} \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}. \quad (3.23)$$

Tal como anteriormente, este ainda não é um intervalo de confiança para a diferença de proporções, uma vez que os extremos do intervalo dependem de  $\pi_1$  e  $\pi_2$ . No entanto, o estimador para a variância de  $\hat{\pi}_1 - \hat{\pi}_2$  é consistente, uma vez que

$$\frac{\hat{\pi}_1(1-\hat{\pi}_1)/n_1 + \hat{\pi}_2(1-\hat{\pi}_2)/n_2}{\pi_1(1-\pi_1)/n_1 + \pi_2(1-\pi_2)/n_2} = \frac{n_2\hat{\pi}_1(1-\hat{\pi}_1) + n_1\hat{\pi}_2(1-\hat{\pi}_2)}{n_2\pi_1(1-\pi_1) + n_1\pi_2(1-\pi_2)} \xrightarrow[n_2 \rightarrow \infty]{n_1 \rightarrow \infty} 1. \quad (3.24)$$

Então, recorrendo uma vez mais, às propriedades dos estimadores para os casos em que estes são consistentes e aplicando o teorema de Slutsky, vem que um intervalo assintótico, com grau de confiança  $1-\alpha$ , para a diferença entre as duas proporções é dado por:

$$\hat{\pi}_1 - \hat{\pi}_2 \pm q_{1-\alpha/2} \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}. \quad (3.25)$$

### 3.2.3 Tabelas de Contingência

Sejam  $X$  e  $Y$  duas variáveis categóricas com  $I$  e  $J$  categorias, respetivamente.

A informação relativa a estas variáveis irá ser representada numa tabela  $I \times J$ , onde as células da tabela irão corresponder às  $IJ$  respostas possíveis. Atribui-se o nome de tabela de contingência quando os valores das células apresentam as frequências de saída de uma amostra. Este termo foi introduzido por Karl Pearson em 1904 (Agresti, 2007).

Estas tabelas podem ser bidimensionais quando as amostras são classificadas sob 2 variáveis ou ainda multidimensional na presença de mais.

Considere-se a seguinte notação:

$$\begin{aligned} y_{ij} & - \text{Valores observados na célula } (i,j), i = 1, \dots, I \text{ e } j = 1, \dots, J \\ n_{i\cdot} & = \sum_{j=1}^J y_{ij} \text{ (total da linha } i), i = 1, \dots, I \\ n_{\cdot j} & = \sum_{i=1}^I y_{ij} \text{ (total da coluna } j), j = 1, \dots, J \\ n & = \sum_{i=1}^I \sum_{j=1}^J y_{ij} \text{ (total de observações da amostra), } i = 1, \dots, I \text{ e } j = 1, \dots, J \end{aligned}$$

Então, a tabela de contingência pode ser apresentada da seguinte forma:

**Tabela 3.2** – Tabela de contingência de dupla entrada  $I \times J$

|          | 1             | 2             | ...      | $j$           | ...      | $J$           | Total        |
|----------|---------------|---------------|----------|---------------|----------|---------------|--------------|
| 1        | $y_{11}$      | $y_{12}$      | ...      | $y_{1j}$      | ...      | $y_{1J}$      | $n_{1\cdot}$ |
| 2        | $y_{21}$      | $y_{22}$      |          | $y_{2j}$      |          | $y_{2J}$      | $n_{2\cdot}$ |
| $\vdots$ | $\vdots$      | $\vdots$      | $\ddots$ | $\vdots$      | $\ddots$ | $\vdots$      | $\vdots$     |
| $i$      | $y_{i1}$      | $y_{i2}$      |          | $y_{ij}$      |          | $y_{iJ}$      | $n_{i\cdot}$ |
| $\vdots$ | $\vdots$      | $\vdots$      | $\ddots$ | $\vdots$      | $\ddots$ | $\vdots$      | $\vdots$     |
| $I$      | $y_{I1}$      | $y_{I2}$      |          | $y_{Ij}$      |          | $y_{IJ}$      | $n_{I\cdot}$ |
| Total    | $n_{\cdot 1}$ | $n_{\cdot 2}$ | ...      | $n_{\cdot j}$ | ...      | $n_{\cdot J}$ | $n$          |

Fonte: Elaboração própria

Seja  $\pi_{ij}$  a probabilidade de  $(X,Y)$  dar-se na célula pertencente à linha  $i$  e coluna  $j$ . Esta probabilidade corresponde à probabilidade conjunta das variáveis  $X$  e  $Y$ .

Na maioria das tabelas de contingência, diz-se que uma variável  $Y$  é a variável resposta e que  $X$  é a variável explicativa. Então, quando se fixa uma categoria de  $X$ ,  $Y$  tem uma distribuição de probabilidade – distribuição condicional -, com probabilidade  $\pi_{j|i}$ , em que  $\sum \pi_{j|i}=1$ .

Este trabalho irá incidir principalmente em tabelas do tipo  $I \times J$ , com  $J=2$ , isto é, tabelas  $I \times 2$ , onde a variável  $Y_i$  assume uma distribuição binomial com parâmetros  $n_i$  e  $\pi_{j|i}$ , cuja função de probabilidade é dada por:

$$P(Y_i = y_{ij}) = \binom{n_i}{y_{ij}} \pi_{j|i}^{y_{ij}} (1 - \pi_{j|i})^{n_i - y_{ij}}. \quad (3.26)$$

### 3.2.3 Teste do Qui-Quadrado

O teste do qui-quadrado é um teste não paramétrico aplicado a dados categóricos para avaliar a discrepância entre as frequências observadas e as frequências esperadas. Então, sendo este um teste não paramétrico não irá depender de parâmetros populacionais, tais como o valor médio ou a variância.

Este é utilizado, essencialmente, para 3 tipos de análises: i) testes de ajustamento (também conhecidos como testes de aderência); ii) testes de homogeneidade; iii) testes de independência.

Embora o teste mais utilizado seja através da estatística de teste de Pearson ( $\chi^2$ ), existem outros testes do qui-quadrado, como por exemplo o teste da razão de verossimilhança.

Estes testes apesar de serem bastante úteis, têm algumas limitações, pois apenas indicam o grau de evidência da associação das diferentes variáveis, não dando assim todas as informações necessárias para um estudo adequado dos dados. Deste modo, para se compreender melhor a natureza das associações, neste trabalho irá recorrer-se posteriormente ao estudo dos resíduos, *odds ratio* e a testes de comparações múltiplas com as respetivas correções dos *p-values*.

Para além desta limitação em relação à informação que disponibilizam, estes também exigem amostras de grandes dimensões e o tratamento das variáveis como nominais, uma vez que é indiferente a ordem das linhas ou colunas no cálculo das estatísticas de teste. Por exemplo, na presença de pelo menos uma variável ordinal ou de uma amostra de pequena dimensão, o teste exato de Fisher revela-se mais apropriado.

#### 3.2.3.1 Pearson ( $\chi^2$ )

É importante referir que tal como qualquer outro teste estatístico, o teste do qui-quadrado também apresenta alguns pressupostos que devem ser respeitados (Agresti, 2002; Berry, 2008):

1. Os grupos têm que ser independentes;
2. As observações têm que ser frequências ou contagens;
3. Cada observação pode pertencer apenas a uma categoria;
4. A amostra deve ser relativamente grande (no máximo, 20% das células pode ter frequência esperada inferior a 5).

A estatística de teste é calculada através da fórmula:

$$\chi^2 = \sum \frac{(\text{frequência observada} - \text{frequência esperada})^2}{\text{frequência esperada}} \quad (3.27)$$

e esta terá (aproximadamente) uma distribuição qui-quadrado com o número de graus de liberdade a variar de acordo com o tipo de teste a fazer.

Quanto maior for o valor desta estatística de teste, maior é a probabilidade de se rejeitar a hipótese nula, ou seja, há pelo menos um grupo de valores observados que não está concordante com os valores esperados, caso a hipótese nula se verificasse.

### 3.2.3.1.1 Teste Qui-Quadrado para Homogeneidade

Considere-se a situação onde são selecionados  $I$  grupos que são classificados de acordo com a variável  $Y$ , podendo assumir dois valores: “sucesso” e “insucesso”. Esta informação pode ser resumida numa tabela de contingência de dupla entrada, como a representada na tabela 3.2.

Assim,  $\pi_{1|i}$  corresponde à probabilidade de sucesso dos indivíduos pertencentes ao grupo  $i$  de  $X$ .

No contexto dos testes do qui-quadrado para a homogeneidade o principal interesse consiste em verificar se as proporções em cada categoria são as mesmas para todos os grupos, isto é, se estas apresentam um comportamento homogêneo. Então, a hipótese nula para este teste é dada por:

$$H_0: \pi_{j|1} = \dots = \pi_{j|I}, j=1, 2. \quad (3.28)$$

Sejam  $e_{ij}$  as frequências esperadas, tal que:

$$e_{ij} = n \pi_{i.} \pi_{.j}. \quad (3.29)$$

Mais uma vez, como estes valores são desconhecidos tem que se recorrer aos seus estimadores.

O estimador para  $\pi$  é dado pelo quociente entre  $Y$  e  $n$ , então pelo mesmo raciocínio utilizado nos testes das proporções, um estimador para  $\pi_{j|i}$  é obtido através do quociente entre  $y_{ij}$  e  $n_{i.}$ .

Assumindo que a hipótese nula é válida e considerando a notação  $\pi_{j|i} = \pi_{.j}$ , tem-se que:

$$\hat{\pi}_j = \hat{\pi}_{j|i} = \frac{\sum_{i=1}^I y_{ij}}{n} = \frac{n_{.j}}{n}, j=1, 2. \quad (3.30)$$

Então, os estimadores de máxima verosimilhança das frequências esperadas são dados por:

$$\hat{e}_{j|i} = n_{i.} \hat{\pi}_j = \frac{n_{i.} n_{.j}}{n}. \quad (3.31)$$

E, substituindo estes valores em (3.27), obtém-se a estatística de teste  $\chi^2$ :

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^2 \frac{(y_{ij} - \hat{e}_{j|i})^2}{\hat{e}_{j|i}}. \quad (3.32)$$

Deste modo, quanto maior a discrepância entre  $y_{ij} - \hat{e}_{j|i}$ , maiores serão os valores produzidos para  $\chi^2$  para valores fixos de  $n$  e, consequentemente, maior será a evidência contra a hipótese da homogeneidade entre os grupos.

Seja  $\chi_0^2$  o valor observado de  $\chi^2$ . O  $p$ -value é dado por  $P(\chi^2 \geq \chi_0^2 / H_0) \simeq P(\chi_{(I-1)(J-1)}^2 \geq \chi_0^2)$ , onde  $\chi_{(I-1)(J-1)}^2$  corresponde à variável aleatória  $\chi^2$  com  $(I-1)(J-1)$  graus de liberdade.

Para amostras de grandes dimensões,  $\chi^2$  tem aproximadamente uma distribuição qui-quadrado com  $(I-1)(J-1)$  graus de liberdade.

### 3.2.3.1.2 Teste Qui-Quadrado para Independência

Considere-se agora o caso em que se seleciona uma determinada amostra de  $n$  indivíduos e estes são classificados segundo as duas variáveis qualitativas  $X$  e  $Y$ .

Num teste de independência, o objetivo é averiguar se as variáveis  $X$  e  $Y$  são independentes, definindo assim as hipóteses:

$$\begin{aligned} H_0: & \text{As variáveis } X \text{ e } Y \text{ são independentes} \\ & \text{vs.} \\ H_1: & \text{As variáveis } X \text{ e } Y \text{ não são independentes.} \end{aligned} \quad (3.33)$$

Por definição, duas variáveis designam-se independentes quando se verifica:

$$P(X=i, Y=j) = P(X=i)P(Y=j), \text{ isto é, } \pi_{ij} = \pi_{i.}\pi_{.j}, \text{ para todo } i=1, \dots, I \text{ e } j=1, \dots, J. \quad (3.34)$$

Recordando a igualdade

$$\pi_{j|i} = \frac{\pi_{ij}}{\pi_{i.}} = \frac{\pi_{i.}\pi_{.j}}{\pi_{i.}} = \pi_{.j}, \text{ para todo } i=1, \dots, I, \quad (3.35)$$

observa-se que cada distribuição condicional de  $Y$  é idêntica à sua distribuição marginal. Assim, duas variáveis são independentes quando  $\pi_{j|1} = \dots = \pi_{j|I}$ , para todo o  $j=1, \dots, J$ .

Tal como no teste anterior, é necessário estimar os valores dos parâmetros para a construção da estatística de teste. Então, os estimadores de máxima verosimilhança para as probabilidades marginais são dados por:

$$\hat{\pi}_{i.} = \frac{n_{i.}}{n} \text{ e } \hat{\pi}_{.j} = \frac{n_{.j}}{n}, \text{ para } i=1, \dots, I \text{ e } j=1, 2. \quad (3.36)$$

E, consequentemente, sob a veracidade de  $H_0$ , tem-se:

$$\hat{\pi}_{ij} = \hat{\pi}_{i.} \times \hat{\pi}_{.j} = \frac{n_{i.}}{n} \times \frac{n_{.j}}{n}, \text{ para } i=1, \dots, I \text{ e } j=1, 2. \quad (3.37)$$

Então, os estimadores das frequências esperadas são dados por:

$$\hat{e}_{ij} = n \times \hat{\pi}_{ij} = n \times \frac{n_{i.} n_{.j}}{n^2} = \frac{n_{i.} n_{.j}}{n}. \quad (3.38)$$

Sendo a estatística de teste para a independência idêntica à do teste para a homogeneidade, conclui-se que esta também se aproxima de uma distribuição qui-quadrado com  $(I-1)(J-1)$  graus de liberdade.

A título de curiosidade, nem sempre foi este o número de graus de liberdade a considerar. De acordo com Agresti (2002), Pearson (1900, 1904, 1922) verificou que ao substituir  $e_{ij}$  pelo seu estimador não iria provocar alterações na distribuição  $\chi^2$  e, como a tabela de contingência apresentava  $I \times J$

categorias, defendia que  $\chi^2$  era assintoticamente uma distribuição  $\chi^2$  com  $(I \times J) - 1$  graus de liberdade. Só mais tarde, R. A. Fisher (1922) verificou que Pearson estava errado, uma vez que como era exigida a estimação das probabilidades marginais, então os graus de liberdade seriam obtidos através de  $(IJ - 1) - (I - 1) - (J - 1) = (I - 1)(J - 1)$ .

### 3.2.3.2 Razão de Verossimilhança ( $G^2$ )

Até agora foram estudados testes construídos em torno da estatística de teste  $\chi^2$ . Todavia existem outros testes do qui-quadrado que se baseiam noutras variáveis, tais como o teste da razão de verossimilhança, também conhecido como teste- $G$ , uma vez que é construído em torno da estatística de teste  $G^2$  (Mcdonald, 2014).

Sendo este uma alternativa ao teste do qui-quadrado de Pearson ( $\chi^2$ ) para a independência, então também deve ser utilizado para variáveis nominais, tal como para amostras de grande dimensão.

Sejam  $L_0$  o valor máximo da função verossimilhança sob  $H_0$  e  $L_1$  o valor máximo generalizado, isto é, sob a validade de  $H_0$  ou de uma hipótese alternativa. Então,  $L_1$  assume valores pelo menos tão grandes quanto  $L_0$ , dado que  $L_0$  resulta sobre o máximo de uma condição para um valor restrito de parâmetros. Deste modo, o rácio ( $\Lambda$ ) obtido através de  $L_0 / L_1$  não pode exceder 1.

Wilks (1935, 1938), citado em Agresti (2007), provou que  $-2 \log \Lambda$  tende para uma distribuição qui-quadrado quando  $n \rightarrow \infty$  com  $(I - 1)(J - 1)$  graus de liberdade, encontrados de forma análoga à estatística de teste  $\chi^2$ .

Então, a estatística de teste  $G^2$  para o teste da razão de verossimilhança é dada por:

$$G^2 = -2 \log \Lambda . \quad (3.39)$$

Uma vez que se pretende testar a independência, sob  $H_0$ , assume-se que

$$\hat{\pi}_{ij} = \hat{\pi}_{i.} \times \hat{\pi}_{.j} = \frac{n_{i.} n_{.j}}{n^2}. \quad (3.40)$$

Enquanto que, num caso geral, tem-se:

$$\hat{\pi}_{ij} = \frac{y_{ij}}{n}. \quad (3.41)$$

Como a razão de verossimilhança é dada por  $\Lambda = L_0 / L_1$ , então:

$$\Lambda = \frac{L_0}{L_1} = \frac{\prod_i \prod_j \left( \frac{n_{i.} n_{.j}}{n^2} \right)^{y_{ij}}}{\prod_i \prod_j \left( \frac{y_{ij}}{n} \right)^{y_{ij}}}. \quad (3.42)$$

Substituindo os valores de (3.42) em (3.39), a estatística de teste é obtida por:

$$G^2 = -2 \log \Lambda = -2 \sum_i \sum_j n_{ij} \log \left( \frac{\hat{\pi}_{ij}}{n_{ij}} \right) = 2 \sum_i \sum_j n_{ij} \log \left( \frac{n_{ij}}{\hat{\pi}_{ij}} \right). \quad (3.43)$$

Então, tal como no caso anterior, quanto maior o valor de  $G^2$ , maior será a evidência contra a independência.

Segundo Mcdonald (2014), o teste- $G$  apresenta resultados idênticos aos obtidos através do teste de Pearson. Todavia, o teste- $G$  pode ser utilizado para estudos mais elaborados e, quanto mais elaborados, maior será a diferença entre estes dois testes, fazendo com que o teste  $G$  seja preferido por muitos, embora o teste de Pearson seja mais familiar entre diversos autores.



Contudo, a convergência para a distribuição qui-quadrado é mais rápida para  $\chi^2$  do que para  $G^2$ , tornando-se fraca para este último caso quando  $n/IJ < 5$ . Enquanto que para  $I$  ou  $J$  grandes, este teste é aceitável para  $\chi^2$  quando algumas frequências esperadas são reduzidas (próximas de 1), mas com a ressalva de que a maioria tem que ser superior a 5 (Agresti, 2002).

### 3.2.4 Odds Ratio

O *Odds Ratio*,  $\theta$ , é uma medida de associação definida pela razão entre duas *Odds*. Este é utilizado para estudar a relação entre duas variáveis nominais em que uma variável corresponde ao “fator” e a segunda ao “acontecimento”, tornando-se assim um dos métodos mais utilizado por diversos autores para a interpretação de tabelas 2x2 devido à sua facilidade face ao teste do qui-quadrado (Cunha, 2014).

Seja  $\pi$  a probabilidade de sucesso, então uma *Odd* é definida por:

$$\Omega = \pi / (1 - \pi). \quad (3.44)$$

Esta assume sempre valores positivos e quando excede 1, conclui-se que há uma maior possibilidade de ocorrer um “sucesso” do que uma “falha”. Por exemplo, se se obtém  $\Omega=2$  diz-se que é mais provável ocorrer duas vezes um “sucesso” do que uma “falha”, isto é, são esperados dois sucessos para cada falha.

Relativamente às tabelas 2x2, a *Odd* de sucesso para a linha  $i$  é  $\Omega_i = \pi_i / (1 - \pi_i)$ , pelo que o *Odds Ratio* entre duas linhas se obtém através de:

$$\theta = \frac{\Omega_1}{\Omega_2} = \frac{\pi_1 / (1 - \pi_1)}{\pi_2 / (1 - \pi_2)}. \quad (3.45)$$

Para a probabilidade conjunta  $\pi_{ij}$ , a definição equivalente de *Odd* para a linha  $i$  é  $\Omega_i = \pi_{i1} / \pi_{i2}$ ,  $i=1,2$ . Então o *Odds Ratio* é dado por:

$$\theta = \frac{\pi_{11} / \pi_{12}}{\pi_{21} / \pi_{22}} = \frac{\pi_{11} \pi_{22}}{\pi_{12} \pi_{21}}. \quad (3.46)$$

Assume-se a independência entre as variáveis  $X$  e  $Y$  quando  $\theta=1$ , enquanto que se  $\theta$  pertencer ao intervalo  $]1, \infty[$ , conclui-se que os indivíduos alusivos à linha/grupo 1 são mais propensos ao sucesso que os da linha/grupo 2, isto é,  $\pi_1 > \pi_2$ . No entanto, é de realçar que, por exemplo, se  $\theta = 3$  não significa que  $\pi_1 = 3\pi_2$ , uma vez que isso corresponde à interpretação do risco relativo<sup>4</sup>! (Agresti, 2002).

Quanto mais distante o valor de  $\theta$  for de 1, maior é a força da associação entre as variáveis. Nos casos em que existe uma célula com probabilidade nula, então  $\theta$  assume valores igual a 0 ou a  $\infty$ .

Note-se que dois valores distintos de  $\theta$  podem representar a mesma associação quando apresentam valores inversos; tal como o valor de  $\theta$  também não sofre alterações quando se trocam as linhas pelas colunas e vice-versa.

Paralelamente, o *Odds Ratio* amostral estima o mesmo parâmetro para cada caso, isto é, para os valores observados  $\{y_{ij}\}$ , o *Odds Ratio* amostral é:

$$(3.47)$$

---

<sup>4</sup> O Risco Relativo corresponde à relação da probabilidade de ocorrer um acontecimento num grupo face ao outro. Geralmente, utilizado para testar a eficácia de alguns tratamentos onde se compara o grupo exposto face ao grupo de controlo (Agresti, 2002)

$$\hat{\theta} = \frac{y_{11} y_{22}}{y_{12} y_{21}}.$$

Atenda-se que embora o *Odds Ratio* seja mais indicado para tabelas 2x2, este também pode ser aplicado em tabelas  $I \times J$ , mas apresenta como desvantagem perder alguma informação.

### 3.2.5 Comparações Múltiplas

Nesta dissertação já foram estudados os casos em que se pretendeu comparar duas amostras. Porém, na prática, é também recorrente encontrar situações onde é necessário fazer várias comparações em simultâneo, surgindo assim os testes para comparações múltiplas que, segundo Holm (1979), é um dos métodos de inferência estatística mais importante nos últimos 50 anos.

Para mais, já foi também analisada a importância de um teste de homogeneidade, onde a hipótese nula refere-se à igualdade das proporções dos diferentes grupos. Contudo, quando esta é rejeitada não se obtém qualquer informação detalhada sobre a população, tornando-se assim interessante comparar os diferentes pares, de modo a perceber quais os grupos que contribuem mais para esta diferença. O mesmo se sucede no caso dos testes de independência, onde reduzidos *p-values* indicam fortes evidências de associação, mas apresentam pouca informação sobre a natureza da mesma.

Anteriormente já foi mencionado que na elaboração de qualquer teste de hipóteses há sempre o risco de se cometerem os erros tipo I ou tipo II, sendo o desejável minimizar essas duas probabilidades. Contudo, conforme se aumenta o número de hipóteses a serem testadas, a probabilidade de se cometer um erro tipo I aumenta também e, conseqüentemente, mais falsos positivos são produzidos, como se demonstra seguidamente:

*Demonstração:* Considerem-se as seguintes hipóteses,  $H_{(1)}, \dots, H_{(n)}$  verdadeiras cujas estatísticas de teste são independentes e que existe pelo menos uma rejeição, isto é, há pelo menos uma ocorrência do erro tipo I. Então:

$$\begin{aligned} P [\text{ocorrer pelo menos 1 erro tipo I}] &= 1 - \prod_{i=1}^n [1 - P (\text{ocorrer uma falsa rejeição})] \\ &= 1 - \prod_{i=1}^n P [\text{não rejeitar } H_{(i)} | H_{(i)} \text{ verdadeira}] = 1 - (1-\alpha)^n. \end{aligned}$$

□

A esta probabilidade (apresentada na demonstração supracitada) é denominada por *Family-Wise Error Rate (FWER)*. Outra medida também importante para este problema é o *False Discovery Rate (FDR)*, que é definido pela proporção esperada de se obter este tipo de erro. (Benjamini *et al*, 1995).

O principal desafio destas comparações múltiplas passa por assegurar que a taxa desse tipo de erro seja mantida ao nível estabelecido *a priori*, garantindo que essa probabilidade não aumenta com o número de testes. Nesse sentido, foram criados vários métodos para controlar a *FWER* ou o *FDR*.

O procedimento de Bonferroni pode ser aplicado a qualquer situação de comparações múltiplas e tem como objetivo controlar a *FWER*, minimizando a presença de falsos positivos. Este é um método de etapa única e, conseqüentemente, origina testes com pouca potência. De modo a contornar este problema foram surgindo novas metodologias denominadas por “melhoramentos de Bonferroni”, que consistem em ter uma pequena proporção de falsos positivos ao invés de eliminá-los, aumentando assim a potência dos testes.

Uma dessas metodologias foi criada por Holm, em 1979, que propôs um método descendente através de múltiplas etapas, ficando conhecido como procedimento de Bonferroni de rejeição sequencial ou procedimento de Holm. Este método comparado com o método tradicional de Bonferroni, já garante uma maior potência. Simes (1986), Hochberg (1988), Hommel (1988) e Rom (1990), mencionados em Benjamini *et al* (1995), foram outros autores que também desenvolveram alguns

procedimentos/modificações de modo a otimizar o procedimento de Bonferroni para o controlo da *FWER*.

Segundo Armstrong (2014), a utilização destes métodos é um tema controverso devido à sua obrigatoriedade, pois existem autores que acreditam não ser necessária qualquer correção, enquanto outros defendem ser fundamental. Neste trabalho serão aplicados os dois primeiros procedimentos mencionados, uma vez que não se tem um elevado número de hipóteses a testar e estes não exigem quaisquer pressupostos relacionados com a independência dos testes.

A maioria destes procedimentos que controlam a *FWER* baseia-se no ajustamento/correção dos *p-values* obtidos nestes testes de comparações múltiplas, onde o *p-value* ajustado para uma hipótese específica corresponde ao menor nível de significância global em que essa hipótese seria rejeitada (Santos, 2013). Uma vez que se trata de procedimentos que envolvem valores obtidos em testes já realizados é-lhes atribuído o nome de procedimentos *post-hoc*.

O procedimento de Marascuilo é uma alternativa a estes testes, uma vez que realiza múltiplas comparações pareadas e tem incorporado um método de correção, dispensando assim o recurso a métodos *post-hoc*.

Nesta secção já foram apresentadas diversas metodologias que têm como objetivo controlar a *FWER*. Porém, este controlo nem sempre é vantajoso. É importante estar ciente das consequências que estão subjacentes a cada tipo de erro, pois em alguns casos é mais importantes assegurar a baixa probabilidade de ocorrer um erro tipo II, tornando-se assim pouco adequada a utilização deste tipo de correções.

### 3.2.5.1 Procedimento de Bonferroni

A correção de Bonferroni, nome atribuído devido ao trabalho do estatístico italiano Carlo Bonferroni (1892-1960), é baseada num método inicialmente proposto por Neyman e Pearson para tomar decisões relativas a estudos que envolviam amostras repetidas (Armstrong *et al*, 2014).

Atualmente, o procedimento de Bonferroni é frequentemente utilizado para contornar o problema do aumento da probabilidade de se cometerem erros tipo I quando se aumenta o número de hipóteses a testar. Deste modo, garante que os *p-values* mantêm o nível de significância  $\alpha$  estabelecido, para todos os testes.

Considerem-se as hipóteses  $H_1, \dots, H_n$  a testar, baseadas nos respetivos *p-values*  $p_1, \dots, p_n$ .

Sejam  $p_{(1)} \leq \dots \leq p_{(n)}$  os *p-values* ordenados e  $H_{(i)}$  a hipótese correspondente ao *p-value*  $p_{(i)}$ .

Então, pelo procedimento clássico de Bonferroni, rejeitam-se todas as hipóteses  $H_{(i)}$ ,  $i=1, \dots, k$ , onde  $k$  corresponde ao maior índice  $i$  que respeita a desigualdade  $p_{(i)} \leq \frac{i}{n} \alpha$ .

**Teorema 3.1** (Benjamini *et al*, 1995) O procedimento de Bonferroni controla a *FWER* para valores menores ou iguais a  $\alpha$ .

*Demonstração:* Suponha-se que existem  $n_0$  hipóteses nulas verdadeiras. Então:

$$FWER = P \{ \text{existir pelo menos um erro tipo I} \} = P \{ \text{rejeitar pelo menos uma } H_i \mid i \in n_0 \} = \\ P \left\{ \bigcup_{i=1}^{n_0} \left( p_{(i)} \leq \frac{\alpha}{n} \right) \right\} \leq \sum_{i=1}^{n_0} \{ P(p_{(i)} \leq \frac{\alpha}{n}) \} = n_0 \times \frac{\alpha}{n} \leq n \times \frac{\alpha}{n} \leq \alpha. ^5$$

□

<sup>5</sup> A desigualdade de Boole afirma que, para qualquer conjunto de acontecimentos  $E_i$ , a probabilidade de ocorrer, pelo menos um, não é maior do que a soma das probabilidades de cada acontecimento, isto é,  $P(\bigcup_i E_i) \leq \sum_i P(E_i)$  (Heyde *et al*, 2001. p.169)

Este método é considerado um teste de etapa única, sendo indicado para pequenas amostras ou poucos grupos a comparar. Consequentemente, como demonstrado no teorema 3.1, torna-se um método conservador, uma vez que a *FWER*, em alguns casos, é muito inferior ao  $\alpha$  estabelecido. Este é também considerado um teste com pouca potência, sendo o ideal nestes testes controlar o erro de tipo I sem baixar demasiado a potência do teste. Então, caso se tenha um vasto número de testes e se pretenda uma maior potência, deve-se recorrer a procedimentos de etapas múltiplas (Shaffer, 1995).

### 3.2.5.2 Procedimento de Holm

Holm (1979) propôs um método que ficou conhecido na literatura como Procedimento de Bonferroni de Rejeição Sequencial ou Procedimento de Holm.

No procedimento clássico de Bonferroni, verificou-se que todos os *p-values* obtidos são comparados com  $\alpha/n$ . Porém, o mesmo não acontece neste procedimento.

No procedimento de Holm recorre-se aos *p-values* obtidos de forma sequencial, tal como o nome sugere, tornando-se assim um método de etapas múltiplas.

Considerem-se novamente as hipóteses a testar  $H_1, \dots, H_n$  e os seus *p-values* associados ordenados  $p_{(1)} \leq \dots \leq p_{(n)}$ , onde  $H_{(i)}$  corresponde à hipótese com o respetivo *p-value*  $p_{(i)}$ .

O procedimento de Holm é constituído pelas seguintes etapas:

i) Se  $p_{(1)} > \alpha/n$ , então as hipóteses  $H_{(1)}, \dots, H_{(n)}$  não são rejeitadas e o procedimento termina. Caso contrário, rejeita-se  $H_{(1)}$  e o teste continua para as restantes  $n-1$  hipóteses ao nível  $\alpha/(n-1)$ . Segue para ii);

ii) Se  $p_{(2)} > \alpha/(n-1)$  não se rejeitam as hipóteses  $H_{(2)}, \dots, H_{(n)}$  e termina. Porém, se  $p_{(1)} \leq \alpha/n$  e  $p_{(2)} \leq \alpha/(n-1)$ , então para além de se rejeitar  $H_{(1)}$ , rejeita-se também  $H_{(2)}$  e repete-se o processo para as  $n-2$  hipóteses, desta vez ao nível  $\alpha/(n-2)$  e, assim sucessivamente.

Em suma, rejeita-se  $H_{(i)}$  sempre que a condição  $p_{(i)} \leq \alpha/(n-i+1)$  se verifica.

De acordo com Holm, ao utilizar este procedimento, os níveis obtidos irão ser comparados com  $\alpha/n, \alpha/(n-1), \dots, \alpha$ , enquanto que no procedimento clássico apenas serão comparados com  $\alpha/n$ , fazendo com que a probabilidade de cometer uma rejeição falsa seja menor ou igual à probabilidade através desta adaptação (Holm, 1979).

Este método pode substituir o clássico, embora o ganho seja pequeno se todas as hipóteses forem “quase verdadeiras”, mas pode ser considerável se as hipóteses forem “completamente falsas” (Holm, 1979).

### 3.2.5.3 Procedimento de Marascuilo

O procedimento de Marascuilo é um método estatístico que tem o intuito de comparar múltiplas proporções par-a-par, permitindo identificar quais as que apresentam diferenças significativas. Este método já incorpora uma correção, não necessitando de recorrer a procedimentos *post-hoc*, como os que foram estudados anteriormente.

Para este procedimento, as hipóteses a estudar são definidas por:

$$H_0: \pi_1 = \dots = \pi_k \text{ vs. } H_1: \exists ij \ i \neq j : \pi_i \neq \pi_j. \quad (3.48)$$

Assumindo que existem  $k$  amostras com dimensão  $n_i$ , onde  $i = 1, \dots, k$ .

O primeiro passo consiste em calcular a diferença entre  $\pi_i$  e  $\pi_j$ , para todo o par tal que  $i \neq j$ . Porém, tal como na secção 3.2.2, os valores de  $\pi_i$  e  $\pi_j$  são desconhecidos, pelo que é necessário recorrer, uma vez mais, aos seus estimadores. Assim, a diferença de proporções para cada par é obtida através de  $\hat{\pi}_i - \hat{\pi}_j$ , para  $i \neq j$ , e o seu valor absoluto irá corresponder à estatística de teste.

O segundo passo, consiste no cálculo dos valores críticos,  $s_{ij}$ , para cada par que podem ser obtidos a partir da fórmula:

$$s_{ij} = \sqrt{\chi^2_{1-\alpha, k-1}} \sqrt{\frac{\hat{\pi}_i(1-\hat{\pi}_i)}{n_i} + \frac{\hat{\pi}_j(1-\hat{\pi}_j)}{n_j}}. \quad (3.49)$$

Posteriormente, estes valores são comparados com as estatísticas de teste correspondentes e se existir pelo menos uma estatística de teste em que o seu valor absoluto apresente valores superiores ao valor crítico respetivo, rejeita-se a hipótese nula e conclui-se que existem diferenças significativas entre as amostras  $i$  e  $j$ .

No cálculo dos valores críticos, ao recorrer-se à distribuição  $\chi^2$  com  $k-1$  graus de liberdade está-se a incorporar a correção dos testes múltiplos (Pereira, 2014).

### 3.2.5.4 Análise de Resíduos

Outro processo que se pode utilizar na tentativa de se identificar os grupos responsáveis por um valor significativo da estatística de teste (isto é, pela rejeição da hipótese nula para a independência) é através dos resíduos padronizados.

Sabe-se que uma comparação célula por célula de frequências observadas e esperadas revela a natureza da evidência sobre a associação. Esta diferença,  $(y_{ij} - e_{ij})$ , é designada por resíduo. Quando este apresenta valores positivos, a frequência observada excede o valor da frequência esperada que a independência prevê.

Geralmente, utilizam-se os resíduos padronizados,  $r_{ij}$ , de modo a seguirem uma distribuição normal de valor médio 0 e desvio-padrão 1 e assim, consegue-se perceber para que valores o resíduo é grande o suficiente para indicar um desvio da independência.

Estes são calculados através da fórmula:

$$r_{ij} = \frac{y_{ij} - e_{ij}}{\sqrt{e_{ij} \left(1 - \frac{n_i}{n}\right) \left(1 - \frac{n_j}{n}\right)}}. \quad (3.50)$$

Posteriormente, se os resíduos padronizados apresentarem valores absolutos superiores ao quantil de probabilidade  $1-\alpha/2$  da distribuição normal-padrão, conclui-se que a célula  $(i,j)$  fornece evidência contra a independência.

Na literatura, muitos autores consideram que  $r_{ij}$  apresentam grandes valores quando são superiores a 2. Este valor deve-se ao quantil de probabilidade 0,975 da normal-padrão, ou seja, 1,96, correspondendo ao valor de  $\alpha=5\%$  (Agresti *et al*, 1997).

## Parte II

---

## Bibliografia

Accenture. (2015). “*Improving Customer Experience is Top Business Priority for Companies Pursuing Digital Transformation, According to Accenture Study*”, consultado em agosto de 2018, através do link: <https://newsroom.accenture.com/news/improving-customer-experience-is-top-business-priority-for-companies-pursuing-digital-transformation-according-to-accenture-study.htm>

Agresti, A. (2002). “*Categorical Data Analysis*”. John Wiley & Sons, Inc., Hoboken, New Jersey, 2<sup>nd</sup> Edition

Agresti, A. (2007). “*An introduction to Categorical Data Analysis*”. Wiley-Interscience. 2<sup>nd</sup> Edition

Agresti, A., Finlay, B. (1997) “*Statistical Methods for the Social Sciences*”, 3<sup>rd</sup> Edition

Antena 1. (2015). “*Eletricidade chegou em força a Portugal há 70 anos*”. Consultado em janeiro de 2018, através do link: [https://www.rtp.pt/noticias/pais/eletricidade-chegou-em-forca-a-portugal-ha-70-anos\\_a798066](https://www.rtp.pt/noticias/pais/eletricidade-chegou-em-forca-a-portugal-ha-70-anos_a798066)

Armstrong, R. A. (2014). “When to use the Bonferroni correction”, *Ophthalmic Physiol Opt* 34:502–508.

Battaglia, D., & Borchardt, M. (2010). “Análise do processo de recuperação de serviços partir das reclamações dos Clientes: estudo de caso em três organizações”. 20: 455-470.

Berry, J. C. (2008). “*An introduction to statistics*”. Mathematics Department, University of Louisiana at Lafayette.

Berry, L., Zeithaml, V., Parasuraman, A. (1985). “Quality counts in services, too”. *Business Horizons*, 28(3): 44-52

Boote, J. (1998). “Towards a comprehensive taxonomy and model of consumer complaining behaviour”. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, 11:140-151.

Crié, D. (2003). “Consumers' complaint behaviour. Taxonomy, typology and determinants: Towards a unified ontology”. *Database Marketing & Customer Strategy Management*, 11: 60-79.

Comissão Europeia, (2013), “*Guia Relativo à aplicação das regras da União Europeia em matéria de auxílios estatais, de «contratos públicos» e de «mercado interno» aos serviços de interesse económico geral e, nomeadamente, aos serviços sociais de interesse geral*”, Bruxelas, SWD (2013)

Contador, J. L., Senne, E. L. F., (2016) “*Testes não paramétricos para pequenas amostras de variáveis não categorizadas: um estudo*”, Gest. Prod., São Carlos

Cunha, A., (2014), “*Modelação Estatística: Estudo do perfil cardiovascular de adultos assintomáticos, da cidade de Benguela*”, Tese de Mestrado, Universidade Aberta, Lisboa

Davidow, M., & Dacin, P. A. (1997). “Understanding and influencing consumer complaint behavior: improving organizational complaint management”. *Advances in Consumer Research*, 24:450-456.

Day, R. L. (1984). “Modeling Choices Among Alternative Responses to Dissatisfaction”. *Advances in Consumer Research*, 11:496-499.

Diário da República, “Decreto-lei nº74/2017, 21 de Junho”

Diez, D. M., Barr, C. D., Çetinkaya-Rundel, M. (2014). *Introductory Statistics with Randomization and Simulation*, 1<sup>st</sup> Edition

EDP. (2016). “*EDP 40 anos: A história recomeça. A energia renova-se*”. Edição Especial

EDP. (2017). “*A nossa visão*”, consultado em agosto de 2018, através do link: <https://portugal.edp.com/pt-pt/a-edp/sobre-nos/a-nossa-visao>

EDP Distribuição. (2017a). “*Relatório de Contas*”

EDP Distribuição. (2017b). “*Relatório de Conformidade*”

EDP Distribuição. (2017c). “*Projetos Piloto para Aperfeiçoamento da Estrutura Tarifária e Introdução de Tarifas Dinâmicas – Consulta Pública da ERSE*”

EGSP. (2016). “Empresas, concessões e alterações relevantes no sector elétrico”, consultado em agosto de 2018, através do link: <http://www.egsp.pt/empresas.php>

ERSE. (2017a). “*Regulamento da Qualidade de Serviço*”, consultado em agosto de 2018, através do link: <http://www.erse.pt/pt/electricidade/regulamentos/qualidadedeservico/Paginas/default.aspx>

ERSE. (2017b). “*A ERSE*”, consultado em janeiro de 2018, através do link: <http://www.erse.pt/pt/aerse/>

Fornell, C., Wernerfelt, B. (1987). “Defensive Marketing Strategy by Customer Complaint Management: A Theoretical Analysis”, *Journal of Marketing Research*, 24(4):337-346.

Heyde, C.C., Seneta, E. (2001). “*Statisticians of the centuries*”, Springer – Verlag New York, Inc.

Holm, S. (1979). “A Simple Sequentially Rejective Multiple Test Procedure”. *Scandinavian Journal Of Statistics*. 6(2):65-70

INE. (2018). “*População residente por local de residência (NUTS - 2013), Sexo e Grupo etário; Anual*”, consultado em agosto de 2018, através do link: [https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine\\_indicadores&indOcorrCod=0008273&ccontexto=bd&selTab=tab2](https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0008273&ccontexto=bd&selTab=tab2)

Invesp. (2017). “*Customer Acquisition vs. Retention Costs – Statistics And Trends*”, consultado em agosto de 2018, através do link: <https://www.invespro.com/blog/customer-acquisition-retention/>



Johnston, R. & Mehra, S. (2002). “Best-practice complaint management. Academy of Management Executive”. 16(4):145-154

Mcdonald, J. H. (2014). “*Handbook of biological statistics*”. Sparky House Publishing, Baltimore. 3<sup>rd</sup> edition

Mckinsey & Company. (2016). “*The CEO guide to customer experience*”, consultado em fevereiro de 2018, através do link: <https://www.mckinsey.com/business-functions/operations/our-insights/the-ceo-guide-to-customer-experience>

Metehan, T., Yasemin, Z. A. (2011). “Demographic Characteristics and Complaint Behavior: An Empirical Study Concerning Turkish Customers”, *International Journal of Business and Social Science*, 2(9)

Monteiro, L. M. C. (2013). “*A aplicação do pensamento lean a um processo de reclamações*”, Tese de mestrado, ISCTE Business School, Lisboa

Parasuraman, A., Zeithaml, V., Berry, L. (1985). “A conceptual model of service quality and its implications for future research”. *Journal of Marketing*. 49(4):41-50

Parasuraman, A., Zeithaml, V., Berry, L. (1988). “SERVQUAL: a multiple item scale for measuring consumer perceptions of service quality”. *Journal of Retailing*. 64(1)

PennState. (2018). “Contingency Tables”, consultado em junho de 2018, através do link: <https://onlinecourses.science.psu.edu/stat414/node/312/>

Pereira, M. (2014). “*Intragenic Initiation in SETD2 Deficiente Cell*”, Tese de Mestrado, Universidade de Lisboa

PORDATA. (2017). “*Poder de compra per capita*”, consultado em agosto de 2018, através do link: “<https://www.pordata.pt/Municipios/Poder+de+compra+per+capita-118>”

PORDATA. (2018a). “*População residente, estimativas a 31 de dezembro: total e por sexo*”, consultado em agosto de 2018, através do link: <https://www.pordata.pt/Municipios/Popula%C3%A7%C3%A3o+residente++estimativas+a+31+de+Dezembro+total+e+por+sexo-136-519>

PORDATA. (2018b). “*População residente com 15 e mais anos sem o ensino secundário (%)*”, consultado em agosto de 2018, através do link: [https://www.pordata.pt/Municipios/Popula%C3%A7%C3%A3o+residente+com+15+e+mais+anos+sem+o+ensino+secund%C3%A1rio+\(percentagem\)-886](https://www.pordata.pt/Municipios/Popula%C3%A7%C3%A3o+residente+com+15+e+mais+anos+sem+o+ensino+secund%C3%A1rio+(percentagem)-886)

Reichheld, F., Markey, R. (2011). “*The Ultime Question 2.0 (Revised and Expanded Edition: How Net Promoter Companies Thrive in a Customer-Driven World)*”. Bain & Company, Boston.

Rollo, M. F. (2015). “Recordar a História da Eletrificação de Portugal em 2015”. *Ingenium* 145:91-93

Santos, D. (2013). “*Comparações múltiplas para dados censurados*”, Tese de mestrado, USP – São Carlos

Shaffer, J. P. (1995). “Multiple Hypothesis Testing”. *Annu. Rev. Psychol.* 46:561-584

Spreng, R., Harrel, G., Mackoy, R. (1995). “Service recovery: Impact on satisfaction and Intentions”. *The Journal of Services Marketing.* 9(1):15-23

Tax, S. S., Brown, S. W. & Chandrashekar, M. (1998). “Customer evaluations of service complaint experiences: Implications for relationship marketing”. *Journal of Marketing.* 62(2):60-76

Benjamini, Y., Hochberg, Y. (1995). “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing”, *Journal of the Royal Statistical Society. Series B (Methodological).* 57(1):289-300

---